(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 0 704 532 A2

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication: 03.04.1996 Bulletin 1996/14

(21) Application number: 95109019.0

(22) Date of filing: 12.06.1995

(51) Int. Cl.[6]: C12N 15/62, C12N 15/12, C07K 14/51, C07K 14/78, C07K 14/47, C07K 14/495, A61K 38/18, A61K 38/39, C12N 1/21
// A61K47/48 , (C12N1/21, C12R1:19)

(84) Designated Contracting States: DE FR GB IT

(30) Priority: 10.06.1994 US 259263

(71) Applicant: United States Surgical Corporation
Norwalk, Connecticut 06856 (US)

(72) Inventors:
- Gruskin, Elliot A.
Killingworth, CT 06419 (US)
- Espino, Pearl
Madison, CT 06443 (US)

(74) Representative: Marsh, Roy David et al
Hoffmann Eitle & Partner
Patent- und Rechtsanwälte
Arabellastrasse 4
D-81925 München (DE)

Remarks:
The applicant has subsequently filed a sequence listing and declared, that it includes no new matter.

(54) Recombinant chimeric proteins and methods of use thereof

(57) A chimeric protein having at least one domain derived from a physiologically active moiety and at least one domain derived from an extracellular matrix protein is provided. Physiologically active domains are derived from physiologically active moieties such as bone morphogenic proteins, transforming growth factors, and dermatan sulfate proteoglycans. The extracellular matrix protein domains are derived from collagen, fibrin, fibrogen, laminins and the like. Recombinant DNA constructs, cloning vectors and transformed cells containing DNA which encodes such chimeric proteins are provided. Methods of using the chimeric proteins, chimeric DNA constructs, cloning vectors containing chimeric DNA construct, and cells transformed with the cloning vectors are also provided. The chimeric proteins can be used as osteogenic agents and/or antiscarring agents.

EP 0 704 532 A2

Description

BACKGROUND

1. Technical Field

Chimeric proteins and more particularly chimeric proteins having a domain which is derived from a physiologically active moiety and a domain derived from an extracellular matrix protein moiety are provided. Further provided are DNA constructs encoding such chimeric proteins and to methods for preparing such chimeric proteins using recombinant DNA technology. Methods for healing tissue including inducing scar reduction and formation of bone and/or cartilage are also provided.

2. Description of Related Art

Chimeric proteins, also known as fusion proteins, are hybrid proteins which combine two or more precursor proteins or peptides through peptide bonds. Fusion proteins may be produced by recombinant technology, i.e., by fusing part of the coding sequence of one gene to the coding sequence of another gene. The fused gene may then be used to transform a suitable organism which then expresses the fusion protein. Such proteins are usually used to test the function of different domains of a protein molecule or to append a locater or binding peptide onto a protein or peptide of interest. For example, portions upstream and partially downstream of human, rat or mouse collagen genes have been fused to other proteins in an attempt to analyze characteristics of transcription. See, e.g., Rossouw, et al. DNA Sequences in the First Intron of the Human ProAlpha-1-I Collagen Gene Enhance Transcription, Journal of Biological Chemistry, 262 (31), pp. 15151-15157 (1987). Genomic imprinting effects have been analyzed by fusing the gene encoding human keratin 18 9 protein with the gene encoding beta-galactosidase (LacZ). See Thorex et al., Parent-Specific Expression of a Human Keratin 18/beta-galactosidase Fusion Gene in Transgenic Mice, Dev. Dyn. (United States), 195 (2) pp. 100-12 (Oct. 1992). European Patent Application 88302039 describes production and purification of a recombinant protein, e.g., collagen, a linker region which may encode a restriction site, and a binding protein for a substrate. The fusion protein is then contacted with a suitable substrate to which it binds and the protein may then be recovered, e.g., from a column.

Extracellular matrix proteins ("EMPs") are found in spaces around or near cells of multicellular organisms and are typically fibrous proteins of two functional types: mainly structural, e.g., collagen and elastin, and mainly adhesive, e.g., fibronectin and laminin. Collagens are a family of fibrous proteins typically secreted by connective tissue cells. Twenty distinct collagen chains have been identified which assemble to form a total of about ten different collagen molecules. A general discussion of collagen is provided by Alberts, et al., The Cell, Garland Publishing, pp. 802-823 (1989), incorporated herein by reference. Other fibrous or filamentous proteins include Type I IF proteins, e.g., keratins; Type II IF proteins, e.g., vimentin, desmin and glial fibrillary acidic protein; Type III IF proteins, e.g., neurofilament proteins; and Type IV IF proteins, e.g., nuclear laminins.

Physiologically active glycoproteins, proteins, peptides and proteoglycans are abundant in living things. Such glycoproteins, proteins, peptides and proteoglycans are involved in a diverse array of cellular or viral functions which include initiation or regulation of metabolism, catabolism, reproduction, growth and repair of various life forms. Physiologically active glycoproteins, proteins, peptides, and proteoglycans include therapeutically active glycoproteins, proteins, peptides, and proteoglycans such as hormones, growth factors, enzymes, ligands and receptors and fragments thereof. Therapeutically active substances include glycoproteins, proteins, peptides and proteoglycans which have been used in medicine and research, e.g., to achieve a beneficial result in relation to disease states, trauma and/or to increase efficiency of normal cellular functions. Examples of therapeutically active glycoproteins, proteins, peptides and proteoglycans include cellular regulatory factors such as interleukins, GCSF, erythropoietin, insulin, growth hormone, ACTH, thyroid hormones, various growth factors, osteogenic or osteoinductive factors, decorin and the like.

Osteogenic agents are any of a family of proteins or peptides that induce formation of bone and/or cartilage. Osteogenin, bone morphogenic protein ("BMP") or osteoinductive protein are other terms which describe proteins having bone inducing activity. BMPs are a family of related proteins that trigger the developmental cascade of bone differentiation by inducing mesenchymal stem cells to grow into a variety of tissues including bone, cartilage, and dentin. The activity of BMPs is particularly useful for repairing large bone defects which may not heal without clinical intervention.

Osteogenic agents have been isolated from demineralized mammalian bone tissue (see, e.g., U.S. Patent Nos. 4,294,753 and 4,761,471). Substantially pure BMPs have been produced by recombinant DNA techniques (see, e.g., U.S. Patent Nos. 5,106,748, 5,187,076, 5,141,905, 5,108,922, 5,166,058, and 5,116,738). U.S. Patent No. 5,168,050 describes the use of a DNA construct having a DNA sequence encoding the precursor portion of BMP-2A ligated to a DNA sequence encoding BMP-2B for obtaining improved expression of BMP-2B.

Certain methods have been employed for inducing formation of bone and/or cartilage with BMPs. When BMP is implanted in viable tissue without a delivery formulation, the BMP resorbs rapidly and does not effectively induce bone formation. Therefore, formulations for delivery or implantation of BMPs have been developed.

The following are examples of attempts to make delivery devices for BMPs. U.S. Patent No. 4,472,840 describes collagen and BMP conjugates or complexes in the form of microporous sponges to induce the formation of osseous tissue in animals or humans. U.S. Patent No. 4,975,527 describes enzyme-solubilized collagen as a carrier of bone morphogenic protein. U.S. Patent No. 4,563,489 describes delivery systems for BMP that are admixtures of biodegradable organic polymers such as polylactic acid and polyglycolic acid.

U.S. Patent No. 5,106,626 describes administration of osteogenic protein extracted from mammalian bone admixed with or absorbed on a matrix such as tricalcium phosphate, hydroxyapatite, thermoplastic polymer materials, collagen, plaster of paris, polylactic acid, polycaprolactic acid, or polyglycolic acid. U.S. Patent Nos. 5,011,691 and 5,250,302 describe methods of purifying osteogenic protein from mammalian bone and combining it with a matrix of porous material such as collagen, homopolymers or copolymers of glycolic acid and lactic acid, hydroxyapatite, or tricalcium phosphate.

It has been suggested that to prevent rapid resorption of BMP from a site of implantation, osteogenic sequestering agents may be used in connection with an admixture of osteogenic protein and a porous polymeric matrix. U.S. Patent No. 5,171,579 describes a composition of an admixture of an osteogenic protein, a porous particulate matrix and an osteogenic protein sequestering amount of blood clot. PCT WO 93/00050 describes an admixture of an osteogenic protein, a polymer matrix of poly (lactic acid), poly (glycolic acid), and copolymers of lactic acid and glycolic acid, and an osteogenic protein-sequestering material which may be alkylcellulose, hyaluronic acid, alginate, poly(ethylene glycol), polyoxyethylene oxide, carboxyvinyl polymer, poly(vinyl alcohol) or carboxymethylcellulose.

Notwithstanding the research done in the area of drug delivery devices, compositions which deliver a clinically effective dose of therapeutic agents over a predetermined period of time to precise target sites that combine easy handling for the medical practitioner with manufacturing convenience are still desirable. Elimination of the above-mentioned separate purified matrix materials, sequestering agents and substitution of more effective therapeutically active compositions would be advantageous.

SUMMARY

Chimeric proteins having a domain derived from at least one extracellular matrix protein and a domain derived from at least one cellular regulatory factor are provided. Suitable domains derived from cell regulatory factors include osteogenic domains, domains derived from a transforming growth factor, and domains derived from dermatan sulfate proteoglycans.

Recombinant DNA constructs having DNA sequences encoding the above mentioned chimeric proteins are provided. Cloning vectors incorporating the above DNA constructs and cells transformed with the vectors are also provided. Therapeutic compositions incorporating the above-mentioned chimeric protein(s) and pharmaceutically acceptable vehicles are provided. For example, a drug delivery composition is provided which has a chimeric protein having a domain derived from a fibrous protein and a domain derived from a physiologically active glycoprotein, protein, peptide and/or proteoglycan.

Methods for preparing a DNA construct including a DNA sequence encoding a cell regulatory factor (such as an osteogenic agent, a transforming growth factor, and/or a dermatan sulfate proteoglycan) operably linked to a DNA sequence encoding an extracellular matrix protein are provided. Also provided are methods of manufacturing osteogenic/extracellular matrix, transforming growth factor/extracellular matrix, and/or dermatan sulfate proteoglycan/extracellular matrix chimeric proteins by transforming a cell with a suitable cloning vector including a DNA construct encoding the osteogenic/extracellular matrix chimeric protein, the transforming growth factor/extracellular matrix chimeric protein, or the dermatan sulfate proteoglycan/extracellular matrix chimeric protein, respectively, culturing the cell in a suitable culture medium and isolating the chimeric protein from the culture medium.

In other embodiments, methods for inducing formation of bone, soft tissue repair, and reducing scar formation involve contacting with a suitable locus an osteogenic chimeric protein, a soft tissue chimeric protein, or an anti-scarring chimeric protein are provided, respectively. Suitable osteogenic chimeric proteins have a domain derived from one or more osteogenic agents and a domain derived from one or more extracellular matrix proteins. Suitable soft tissue chimeric proteins have a domain derived from at least one transforming growth factor and a domain derived from one or more extracellular matrix proteins. Suitable anti-scarring chimeric proteins have a domain derived from dermatan sulfate proteoglycan and a domain derived from one or more extracellular matrix proteins. Further provided are methods for inducing bone formation, soft tissue repair, and reducing scar formation by contacting the osteogenic chimeric protein, the soft tissue chimeric protein, or the anti-scarring chimeric protein, respectively, with an implant at a suitable locus in viable tissue.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 depicts a nucleic acid sequence which encodes a BMP2B/collagen IA protein construct.
Fig. 2 depicts a nucleic acid sequence which encodes a transforming growth factor β_1/collagen IA protein construct.
Fig. 3 depicts a nucleic acid sequence which encodes a dermatan sulfate proteoglycan/collagen IA protein construct.

Fig. 4 depicts a nucleic acid sequence which encodes a dermatan sulfate proteoglycan peptide/collagen IA protein construct.

Fig. 5 depicts an amino acid sequence for a BMP2B/collagen IA chimeric protein.

Fig. 6 depicts an amino acid sequence for a TGFβ/collagen IA chimeric protein.

Fig. 7 depicts an amino acid sequence for a dermatan sulfate proteoglycan/collagen IA chimeric protein.

Fig. 8 depicts an amino acid sequence for a dermatan sulfate proteoglycan peptide/collagen IA chimeric protein.

Fig. 9 depicts a pMal cloning vector containing a polylinker cloning site.

Fig. 10 depicts a polylinker cloning site contained in a pMal cloning vector of Fig. 9.

Fig. 11 depicts a pMal cloning vector containing a BMP2B/collagen IA DNA construct.

Fig. 12 depicts a pMal cloning vector containing a TGFβ/collagen IA DNA construct.

Fig. 13 depicts a pMal cloning vector containing a decorin/collagen IA DNA construct.

Fig. 14 depicts a pMal cloning vector containing a decorin peptide/collagen IA DNA construct.

DETAILED DESCRIPTION

Chimeric proteins provide an integrated combination of a therapeutically active domain containing one or more therapeutically active moieties and an extracellular matrix protein domain containing one or more EMP moieties. The EMP domain provides an integral vehicle for delivery of the therapeutically active moiety to a target site. The two domains are linked covalently by one or more peptide bonds contained in a linker region. As used herein, integrated or integral means characteristics which result from the covalent association of one or more domains of the inventive chimeric proteins. The therapeutically active moieties disclosed herein are typically made of amino acids linked to form peptides, proteins, glycoproteins or proteoglycans.

The inherent characteristics of EMPs are ideal for use as a vehicle for the therapeutic moiety. Examples of suitable EMPs are collagen, elastin, fibronectin, and fibrin. Fibrillar collagens (Type I, II and III) assemble into ordered polymers and often aggregate into larger bundles. Type IV collagen assembles into sheetlike meshworks. Elastin molecules form filaments and sheets in which the elastin molecules are highly cross-linked to one another and provides good elasticity and high tensile strength. The cross-linked, random-coiled structure of the fiber network allows it to stretch and recoil like a rubber band. Fibronectin is a large fibril forming glycoprotein, which, in one of its forms, consists of highly insoluble fibrils cross-linked to each other by disulfide bonds. Fibrin is an insoluble protein formed from fibrinogen by the proteolytic activity of thrombin during the normal clotting of blood.

The molecular and macromolecular morphology of the above EMPs defines networks or matrices to provide substratum or scaffolding in integral covalent association with the therapeutically active moiety. The networks or matrices formed by the EMP domain provide an environment particularly well suited for ingrowth of autologous cells involved in growth, repair and replacement of existing tissue. The integral therapeutically active moieties covalently bound within the networks or matrices provide maximum exposure of the active agents to their targets to elicit a desired response.

Implants formed of or from the present chimeric proteins provide sustained release activity in or at a desired locus or target site. Unlike the above-described compositions discussed in the Background which incorporate a vehicle not covalently linked to an EMP, the therapeutically active domain of the present chimeric protein is not free to separately diffuse or otherwise be transported away from the vehicle which carries it, absent cleavage of peptide bonds. Consequently, chimeric proteins provide an effective anchor for therapeutic activity which allows the activity to be confined a target location for a prolonged duration. Because the supply of therapeutically active agent does not have to be replenished as often, smaller amounts of therapeutically active agent may be used over the course of therapy. Consequently, certain advantages provided by the inventive chimeric proteins are a decrease or elimination of local and systemic side effects, less potentiation or reduction in therapeutic activity with chronic use, and minimization of drug accumulation in body tissues with chronic dosing.

Use of recombinant technology allows manufacturing of nonimmunogenic chimeric proteins. The DNA encoding both the therapeutically active moiety and EMP moiety should preferably be derived from the same species as the patient being treated to avoid an immunogenic reaction. For example, if the patient is human, the therapeutically active moiety as well as the EMP moiety is preferably derived from human DNA.

Osteogenic/EMP chimeric proteins provide biodegradable and biocompatible agents for inducing bone formation at a desired site. In one embodiment a BMP moiety is covalently linked with an EMP to form a chimeric protein. The BMP moiety induces osteogenesis and the extracellular matrix protein moiety provides an integral substratum or scaffolding for the BMP moiety and cells which are involved in reconstruction and growth. Compositions containing the BMP/EMP chimeric protein provide effective sustained release delivery of the BMP moiety to desired target sites. The method of manufacturing such an osteogenic agent is efficient because the need for extra time consuming steps such as purifying EMP and then admixing it with the purified BMP are eliminated. An added advantage of the BMP/EMP chimeric protein results from the stability created by the covalent bond between BMP and the EMP, i.e., the BMP portion is not free to separately diffuse away from the EMP, thus providing a more stable therapeutic agent.

Bone morphogenic proteins are class identified as BMP-1 through BMP-9. A preferred osteogenic protein for use in human patients is human BMP-2B. A BMP-2B/collagen IA chimeric protein is illustrated in Fig. 5. The protein sequence illustrated in Fig. 5 includes a collagen helical domain depicted at amino acids 1-1057 and a mature form of BMP2B at amino acids 1060-1169. The physical properties of the chimeric protein are dominated in part by the EMP component. In the case of a collagen moiety, a concentrated solution of chimeric protein will have a gelatinous consistency that allows easy handling by the medical practitioner. The EMP moiety acts as a sequestering agent to prevent rapid desorption of the BMP moiety from the desired site and provide sustained release of BMP activity. As a results the BMP moiety remains at the desired site for a period of time necessary to effectively induce bone formation. The EMP moiety also provides a matrix which allows a patient's autologous cells, e.g., chondrocytes and the like, which are normally involved in osteogenesis to collect therein and form an autologous network for new tissue growth. The gelatinous consistency of the chimeric protein also provides a useful and convenient therapeutic manner for immobilizing active BMP on a suitable vehicle or implant for delivering the BMP moiety to a site where bone growth is desired.

The BMP moiety and the EMP moiety are optionally linked together by linker sequences of amino acids. Examples of linker sequences used are illustrated within the sequences depicted in Figs. 1-4 and described in more detail below. Linker sequences may be chosen based on particular properties which they impart to the chimeric protein. For example, amino acid sequences such as Ile-Glu-Gly-Arg and Leu-Val-Pro-Arg are cleaved by Factor Xa and Thrombin enzymes, respectively. Incorporating sequences which are cleaved by proteolytic enzymes into chimeric proteins provides cleavage at the linker site upon exposure to the appropriate enzyme and separation of the two domains into separate entities. It is contemplated that numerous linker sequences can be incorporated into any of the chimeric proteins.

In another embodiment, a chimeric DNA construct includes a gene encoding an osteogenic protein or a fragment thereof linked to a gene encoding an EMP or a fragment thereof. The gene sequences for various BMPs are known, see, e.g., U.S. Patent Nos. 4,294,753, 4,761,471, 5,106,748, 5,187,076, 5,141,905, 5,108,922, 5,166,058, 5,116,738 and 5,168,050, each incorporated herein by reference. A BMP-2B gene for use with this invention is synthesized by ligating oligonucleotides encoding a BMP protein. The oligonucleotides encoding BMP-2B are synthesized using an automated DNA synthesizer (Beckmen Oligo-1000). In a preferred embodiment, the nucleotide sequence encoding the BMP is maximized for expression in E. coli. This is accomplished by using E. coli utilization tables to translate the sequence of amino acids of the BMP into codons that are utilized most often by E. coli. Alternatively, native DNA encoding BMP isolated from mammals including humans may be purified and used.

The BMP gene and the DNA sequence encoding an extracellular matrix protein are cloned by standard genetic engineering methods as described in Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor 1982, hereby incorporated by reference.

The DNA sequence corresponding to the helical region of collagen I(a) is cloned from a human fibroblast cell line. Two sets of polymerase chain reactions are carried out using cDNA prepared by standard methods from AG02261A cells. The first pair of PCR primers include a 5' primer bearing an XmnI linker sequence and a 3' primer bearing the BsmI site at nucleotide number 1722. The resulting PCR product consists of sequence from position 1 to 1722. The second pair of primers includes the BsmI site at 1722 and a linker sequence at the 3' end bearing a BglII site. The resulting PCR products consists of sequence from position 1722 to 3196. The complete helical sequence is assembled by standard cloning techniques. The two PCR products are ligated together at the BsmI site, and the combined clone is inserted into any vector with XmnIBglII sites of XmnI-BamHI sites such as pMALc2-vector.

To clone the BMP-2B gene, total cellular RNA is isolated from human osteosarcoma cells (U-20S) by the method described by Robert E. Farrel Jr. (Academic Press, CA, 1993 pp.68-69) (herein incorporated by reference). The integrity of the RNA is verified by spectrophotometric analysis and electrophoresis through agarose gels. Typical yields of total RNA are 50 µg from a 100mm confluent tissue culture dish. The RNA is used to generate cDNA by reverse transcription using the Superscript pre-amplification system by Gibco BRL. The cDNA is used as template for PCR amplification using upstream and downstream primers specific for BMP-2B (GenBank HUMBMP2B accession # M22490). The resulting PCR product consists of BMP-2B sequence Eom position 1289-1619. The PCR product is resolved by electrophoresis through agarose gels, purified with gene clean (BIO 101) and ligated into pMal-c2 vector (New England Biolabs). The helical domain of human collagen I(a) chain is cloned in a similar manner. However, the total cellular RNA is isolated from a human fibroblast cell line (AG02261A human skin fibroblasts).

A chimeric BMP/EMP DNA construct is obtained by ligating a synthetic BMP gene to a DNA sequence encoding an EMP such as collagen, fibrin, fibronectin, elastin or laminin. However, the invention is not limited to these particular proteins. Fig. 1 illustrates a DNA construct which encodes a BMP-2B/collagen IA chimeric protein. The coding sequence for an EMP may be ligated upstream and/or downstream and in-frame with a coding sequence for the BMP. The DNA encoding an EMP may be a portion of the gene or an entire EMP gene. Furthermore, two different EMPs may be ligated upstream and downstream from the BMP.

The BMP-2B/collagen IA chimeric protein illustrated in Fig. 1 includes an XmnI linker sequence at base pairs (bp) 1-19, a collagen helical domain (bp 20-3190), a BglII/BamHI linker sequence (bp 3191-3196), a mature form of BMP-2B (bp 3197-3529) and a HindIII linker sequence (bp 3530-3535).

Any combination of growth factor and matrix protein sequences are contemplated including repeating units, or multiple arrays of each segment in any order. Incorporation of fragments of both matrix and growth factor proteins is also contemplated. For example, in the case of collagen, only the helical domain may be included. Other matrix proteins have defined domains, such as laminin, which has EGF-like domains. In these cases, specific functionalities can be chosen to achieve desired effects. Moreover, it may be useful to combine domains from disparate matrix proteins, such as the helical region of collagen and the cell attachment regions of fibronectin. In the case of growth factors, specific segments have been shown to be removed from the mature protein by post translational processing. Chimeric proteins can be designed to include only the mature biologically active region. For example, in the case of BMP-2B only the final 110 amino acids are found in the active protein.

In another embodiment, a transforming growth factor (TGF) moiety is covalently linked with an EMP to form a chimeric protein. The TGF moiety increases efficacy of the body's normal soft tissue repair response and also induces osteogenesis. Consequently, TGF/EMP chimeric proteins may be used for either or both functions. One of the fundamental properties of the TGFβs is their ability to turn on various activities that result in the synthesis of new connective tissue. See, Piez and Sporn eds., Transforming Growth Factor-βs Chemistry, Biology and Therapeutics, Annals of the New York Academy of Sciences, Vol. 593, (1990). TGF-β is known to exist in at least five different isoforms. The DNA sequence for Human TGF-β_1 is known and has been cloned. See Derynck et al., Human Transforming Growth Factor-Beta cDNA Sequence and Expression in Tumour Cell Lines, Nature, Vol. 316, pp. 701-705 (1985), herein incorporated by reference. TGF-β_2 has been isolated from bovine bone, human glioblastoma cells and porcine platelets. TGF-β_3 has also been cloned. See ten Dijke, et al., Identification of a New Member of the Transforming Growth Factor-β Gene Family, Proc. Natl. Acad. Sci. (USA), Vol. 85, pp. 4715-4719 (1988) herein incorporated by reference.

A TGF-β/EMP chimeric protein incorporates the known activities of TGF-βs and provides integral scaffolding or substratum of the EMP as described above to yield a composition which further provides sustained release focal delivery at target sites.

The TGF-β moiety and the EMP moiety are optionally linked together by linker sequences of amino acids. Linker sequences may be chosen based upon particular properties which they impart to the chimeric protein. For example, amino acid sequences such as Ile-Glu-Glyn-Arg and Leu-Val-Pro-Arg are cleaved by Factor Xa and Thrombin enzymes, respectively. Incorporating sequences which are cleaved by proteolytic enzymes into the chimeric protein provides cleavage at the linker site upon exposure to the appropriate enzyme and separation of the domains into separate entities. Fig. 6 depicts an amino acid sequence for a TGF-β_I/collagen IA chimeric protein. The illustrated amino acid sequence includes the collagen helical domain (1-1057) and a mature form of TGF-β_I (1060-1171).

A chimeric DNA construct includes a gene encoding TGF-β_I or a fragment thereof, or a gene encoding TGF-β_2 or a fragment thereof, or a gene encoding TGF-β_3 or a fragment thereof, ligated to a DNA sequence encoding an EMP protein such as collagen (I-IV), fibrin, fibronectin, elastin or laminin. A preferred chimeric DNA construct combines DNA encoding TGF-β_I, a DNA linker sequence, and DNA encoding collagen IA. A chimeric DNA construct containing TGF-β_I gene and a collagen IA gene is shown in Fig. 2. The illustrated construct includes an XmnI linker sequence (bp 1-19), DNA encoding a collagen helical domain (bp 20-3190), a BglII linker sequence (bp 3191-3196), DNA encoding a mature form of TGF-β_I (3197-3535), and an XbaI linker sequence (bp 3536-3541).

The coding sequence for EMP may be ligated upstream and/or downstream and in-frame with a coding sequence for the TGFβ. The DNA encoding the extracellular matrix protein may encode a portion of fragment of the EMP or may encode the entire EMP. Likewise, the DNA encoding the TGF-β may be one or more fragments thereof or the entire gene. Furthermore, two or more different TGF-βs or two or more different EMPs may be ligated upstream or downstream of alternate moieties.

In yet another embodiment, a dermatan sulfate proteoglycan moiety, also known as decorin or proteoglycan II, is covalently linked with an EMP to form a chimeric protein. Decorin is known to bind to type I collagen and thus affect fibril formation, and to inhibit the cell attachment promoting activity of collagen and fibrinogen by binding to such molecules near their cell binding sites. Chimeric proteins which contain a decorin moiety act to reduce scarring of healing tissue. The primary structure of the core protein of decorin has been deduced from cloned cDNA. See Krusius et al., Primary Structure of an Extracellular Matrix Proteoglycan Core Protein-Deduced from Cloned cDNA, Proc. Natl. Acad. Sci. (USA), Vol. 83, pp. 7683-7687 (1986) incorporated herein by reference.

A decorin/EMP chimeric protein incorporates the known activities of decorin and provides integral scaffolding or substratum of the EMP as described above to yield a composition which allows sustained release focal delivery to target sites. Fig. 7 illustrates a decorin/collagen IA chimeric protein in which the collagen helical domain includes amino acids 1-1057 and the TGF-β mature protein includes amino acids 1060- 1171. Fig. 8 illustrates a decorin peptide/collagen IA chimeric protein in which the collagen helical domain includes amino acids 1- 1057 and the decorin peptide fragment includes amino acids 1060- 1107. The decorin peptide fragment is composed of P46 to G93 of the mature form of decorin.

Further provided is a chimeric DNA construct which includes a gene encoding decorin or one or more fragments thereof, optionally ligated via a DNA linker sequence to a DNA sequence encoding an EMP such as collagen (I-IV), fibrin, fibronectin, elastin or laminin. A preferred chimeric DNA construct combines DNA encoding decorin, a DNA linker sequence, and DNA encoding collagen IA. A chimeric DNA construct containing a decorin gene and a collagen IA gene

is shown in Fig. 3. The illustrated construct includes an XmnI linker sequence (bp 1-19), DNA encoding a collagen helical domain (bp 20-3190), a BglII linker sequence (bp 3191-3196), DNA encoding a mature form of decorin (bp 3197-4186) and a PstI linker sequence. A chimeric DNA construct containing a decorin peptide gene and a collagen IA gene is shown in Fig. 4. The illustrated construct includes an XmnI linker sequence (bp 1- 19), DNA encoding a collagen helical domain (bp 20-3190), a BglII linker sequence (bp 3191-3196), DNA encoding a peptide fragment of decorin (bp 3197-3343), and a PstI linker sequence (bp 3344-3349).

The coding sequence for an EMP may be ligated upstream and/or downstream and in-frame with a coding sequence for decorin. The DNA encoding the EMP may encode a portion or fragment of the EMP or may encode the entire EMP. Likewise, the DNA encoding decorin may be a fragment thereof or the entire gene. Furthermore, two or more different EMPs may be ligated upstream from the DNA encoding decorin moiety.

Any of the above described chimeric DNA constructs may be incorporated into a suitable cloning vector. Fig. 9 depicts a pMal cloning vector containing a polylinker cloning site. Preferred cloning vectors are the plasmids pMal-p2 and pMal-c2 (commercially available from New England Biolabs). The desired chimeric DNA construct is incorporated into a polylinker sequence of the plasmid which contains certain useful restriction endonuclease sites which are depicted in Fig. 10. The pMal-p2 polylinker sequence has XmnI, EcoRI, BamHI, HindIII, XbaI, SalI and PstI restriction endonuclease sites which are depicted in Fig. 10. The polylinker sequence is digested with an appropriate restriction endonuclease and the chimeric construct is incorporated into the cloning vector by ligating it to the DNA sequences of the plasmid. The chimeric DNA construct may be joined to the plasmid by digesting the ends of the DNA construct and the plasmid with the same restriction endonuclease to generate "sticky ends" having 5' phosphate and 3' hydroxyl groups which allow the DNA construct to anneal to the cloning vector. Gaps between the inserted DNA construct and the plasmid are then sealed with DNA ligase. Other techniques for incorporating the DNA construct into plasmid DNA include blunt end ligation, poly(dA.dT)tailing techniques, and the use of chemically synthesized linkers. An alternative method for introducing the chimeric DNA construct into a cloning vector is to incorporate the DNA encoding the extracellular matrix protein into a cloning vector already containing a gene encoding a therapeutically active moiety.

The cloning sites in the above-identified polylinker site allow the cDNA for the collagen IA/BMP-2B chimeric protein illustrated in Fig. 1 to be inserted between the XmnI and the HindIII sites. The cDNA encoding the collagen LtTGF-β_l protein illustrated in Fig. 2 is inserted between the XmnI and the XbaI sites. The cDNA encoding the collagen IA/decorin protein illustrated in Fig. 3 is inserted between the XmnI and the PstI sites. The cDNA encoding the collagen IA/decorin peptide (dec 1) illustrated in Fig. 4 is inserted between the XmnI and PstI sites.

Plasmids containing the chimeric DNA construct are identified by standard techniques such as gel electrophoresis. Procedures and materials for preparation of recombinant vectors, transformation of host cells with the vectors, and host cell expression of polypeptides are described in Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor 1982 hereby incorporated by reference. Generally, prokaryotic or eukaryotic host cells may be transformed with the recombinant DNA plasmids. Transformed host cells may be located through phenotypic selection genes of the cloning vector which provide resistance to a particular antibiotic when the host cells are grown in a culture medium containing that antibiotic.

Transformed host cells are isolated and cultured to promote expression of the chimeric protein. The chimeric protein may then be isolated from the culture medium and purified by various methods such as dialysis, density gradient centrifugation, liquid column chromatography, isoelectric precipitation, solvent fractionation, and electrophoresis. However, purification of the chimeric protein by affinity chromatography is preferred whereby the chimeric protein is purified by ligating it to a binding protein and contacting it with a ligand or substrate to which the binding protein has a specific affinity.

In order to obtain more effective expression of mammalian or human eukaryotic genes in bacteria (prokaryotes), the mammalian or human gene should be placed under the control of a bacterial promoter. A protein fusion and purification system is employed to obtain the chimeric protein. Preferably, any of the above-described chimeric DNA constructs is cloned into a pMal vector at a site in the vector's polylinker sequence. As a result, the chimeric DNA construct is operably fused with the malE gene of the pMal vector. The malE gene encodes maltose binding protein (MBP). Fig. 11 depicts a pMal cloning vector containing a BMP/collagen DNA construct. A spacer sequence coding for 10 asparagine residues is located between the malE sequence and the polylinker sequence. This spacer sequence insulates MBP from the protein of interest. Figs. 12, 13 and 14 depict pMal cloning vectors containing DNA encoding TGF-β1, decorin and a decorin peptide, respectively. The pMal vector containing any of the chimeric DNA constructs fused to the malE gene is transformed into E. coli. This technique utilizes the PtaC promoter of the malE gene.

The E. coli is cultured in a medium which induces the bacteria to produce the maltose binding protein fused to the chimeric protein. The MBP contains a 26 amino acid N-terminal signal sequence which directs the MBP-chimeric protein through the E. coli cytoplasmic membrane. The protein can then be purified from the periplasm. Alternatively, the pMal-c2 cloning vector can be used with this protein fusion and purification system. The pMal-c2 vector contains an exact deletion of the malE signal sequence which results in cytoplasmic expression of the fusion protein. A crude cell extract containing the fusion protein is prepared and poured over a column of amylose resin. Since MBP has an affinity for the amylose it binds to the resin. Alternatively, the column can include any substrate for which MBP has a specific affinity. Unwanted proteins present in the crude extract are washed through the column. The MBP fused to the chimeric protein

is eluted from the column with a neutral buffer containing maltose or other dilute solution of a desorbing agent for displacing the hybrid polypeptide. The purified MBPchimeric protein is cleaved with a protease such as factor Xa protease to cleave the MBP from the chimeric protein. The pMal-p2 plasmid has a sequence encoding the recognition site for protease factor Xa which cleaves after the amino acid sequence Isoleucine-Glutamic acid-Glycine-Arginine of the polylinker sequence.

The chimeric protein is then separated from the cleaved MBP by passing the mixture over an amylose column. An alternative method for separating the MBP from the chimeric protein is by ion exchange chromatography. This system yields up to 100mg of MBP-chimeric protein per liter of culture. See Riggs, P., in Ausebel, F.M., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K. (eds.) Current Protocols in Molecular Biology, Supplement 19 (16.6.116.6.10) (1990) Green Associates/Wiley Interscience, New York, New England Biolabs (cat # 800-65S 9pMALc2) pMal protein fusion and purification system hereby incorporated by reference. (See also European Patent No. 286 239 herein incorporated by reference which discloses a similar method for production and purification of a protein such as collagen.)

Other protein fusion and purification systems may be employed to produce chimeric proteins. Prokaryotes such as E. coli are the preferred host cells for expression of the chimeric protein. However, systems which utilize eukaryote host cell lines are also acceptable such as yeast, human, mouse, rat, hamster, monkey, amphibian, insect, algae, and plant cell lines. For example, HeLa (human epithelial), 3T3 (mouse fibroblast), CHO (Chinese hamster ovary), and SP 2 (mouse plasma cell) are acceptable cell lines. The particular host cells that are chosen should be compatible with the particular cloning vector that is chosen.

Another acceptable protein expression system is the Baculovirus Expression System manufactured by Invitrogen of San Diego, California. Baculoviruses form prominent crystal occlusions within the nuclei of cells they infect. Each crystal occlusion consists of numerous virus particles enveloped in a protein called polyhedrin. In the baculovirus expression system, the native gene encoding polyhedrin is substituted with a DNA construct encoding a protein or peptide having a desired activity. The virus then produces large amounts of protein encoded by the foreign DNA construct. The preferred cloning vector for use with this system is pBlueBac III (obtained from Invitrogen of San Diego, California). The baculovirus system utilizes the *Autograph californica* multiple nuclear polyhidrosis virus (AcMNPV) regulated polyhedrin promoter to drive expression of foreign genes. AcMNPV is isolated from a moth called the California looper. The chimeric gene, i.e., the DNA construct encoding the chimeric protein, is inserted into the pBlueBac III vector immediately downstream from the baculovirus polyhedrin promoter.

The pBlueBac III transfer vector contains a B-galactosidase reporter gene which allows for identification of recombinant virus. The B-galactosidase gene is driven by the baculovirus ETL promoter (PETL) which is positioned in opposite orientation to the polyhedrin promoter (PpH) and the multiple cloning site of the vector. Therefore, recombinant virus coexpresses B-galactosidase and the chimeric gene.

Spodoptera frugeperda (Sf9) insect cells are then cotransfected with wild type viral DNA and the pBlueBac III vector containing the chimeric gene. Recombination sequences in the pBlueBac III vector direct the vector's integration into the genome of the wild type baculovirus. Homologous recombination occurs resulting in replacement of the native polyhedrin gene of the baculovirus with the DNA constuct encoding the chimeric protein. Wild type baculovirus which do not contain foreign DNA express the polyhedrin protein in the nuclei of the infected insect cells. However, the recombinants do not produce polyhedrin protein and do not produce viral occlusions. Instead, the recombinants produce the chimeric protein.

Alternative insect host cells for use with this expression system are Sf21 cell line derived from *Spodoptera frugeperda* and High Five cell lines derived from *Trichoplusia ni*.

Other acceptable cloning vectors include phages, cosmids or artificial chromosomes. For example, bacteriophage lambda is a useful cloning vector. This phage can accept pieces of foreign DNA up to about 20,000 base pairs in length. The lambda phage genome is a linear double stranded DNA molecule with single stranded complementary (cohesive) ends which can hybridize with each other when inside an infected host cell. The lambda DNA is cut with a restriction endonuclease and the foreign DNA, e.g. the DNA to be cloned, is ligated to the phage DNA fragments. The resulting recombinant molecule is then packaged into infective phage particles. Host cells are infected with the phage particles containing the recombinant DNA. The phage DNA replicates in the host cell to produce many copies of the desired DNA sequence.

Cosmids are hybrid plasmid/bacteriophage vectors which can be used to clone DNA fragments of about 40,000 base pairs. Cosmids are plasmids which have one or more DNA sequences called "cos" sites derived from bacteriophage lambda for packaging lambda DNA into infective phage particles. Two cosmids are ligated to the DNA to be cloned. The resulting molecule is packaged into infective lambda phage particles and transfected into bacteria host cells. When the cosmids are inside the host cell they behave like plasmids and multiply under the control of a plasmid origin of replication. The origin of replication is a sequence of DNA which allows a plasmid to multiply within a host cell.

Yeast artificial chromosome vectors are similar to plasmids but allow for the incorporation of much larger DNA sequences of about 400,000 base pairs. The yeast artificial chromosomes contain sequences for replication in yeast. The yeast artificial chromosome containing the DNA to be cloned is transformed into yeast cells where it replicates thereby producing many copies of the desired DNA sequence. Where phage, cosmids, or yeast artificial chromosomes

are employed as cloning vectors, expression of the chimeric protein may be obtained by culturing host cells that have been transfected or transformed with the cloning vector in a suitable culture medium.

Chimeric proteins disclosed herein are intended for use in treating mammals or other animals. The therapeutically active moieties described above, namely, osteogenic agents such as BMPs, TGFs, decorin, and/or fragments of each of them, are all to be considered as being or having been derived from cellular regulatory factors for purposes. The chimeric proteins and DNA constructs which incorporate a domain derived from one or more cellular regulatory factors can be used for in vivo therapeutic treatment, in vitro research or for diagnostic purposes in general

When used in vivo, formulations containing the inventive chimeric proteins may be placed in direct contact with viable tissue, including bone, to induce or enhance growth, repair and/or replacement of such tissue. This may be accomplished by applying a chimeric protein directly to a target site during surgery. It is contemplated that minimally invasive techniques such as endoscopy are to be used to apply a chimeric protein to a desired location. Formulations containing the chimeric proteins disclosed herein may consist solely of one or more chimeric proteins or may also incorporate one or more pharmaceutically acceptable adjuvants.

In an alternate embodiment, any of the above-described chimeric proteins may be contacted with, adhered to, or otherwise incorporated into an implant such as a drug delivery device or a prosthetic device. Chimeric proteins may be microencapsulated or macroencapsulated by liposomes or other membrane forming materials such as alginic acid derivatives prior to implantation and then implanted in the form of a pouchlike implant. The chimeric protein may be microencapsulated in structures in the form of spheres, aggregates of core material embedded in a continuum of wall material or capillary designs. Microencapsulation techniques are well known in the art and are described in the Encyclopedia of Polymer Science and Engineering, Vol. 9, pp. 724 et seq. (1980) hereby incorporated by reference.
Chimeric proteins may also be coated on or incorporated into medically useful materials such as meshes, pads, felts, dressings or prosthetic devices such as rods pins, bone plates, artificial joints, artificial limbs or bone augmentation implants. The implants may, in part, be made of biocompatable materials such as glass, metal, ceramic, calcium phosphate or calcium carbonate based materials. Implants having biocompatible biomaterials are well known in the art and are all suitable for use. Implant biomaterials derived from natural sources such as protein fibers, polysaccharides, and treated naturally derived tissues are described in the Encyclopedia of Polymer Science and Engineering, Vol. 2, pp. 267 et seq. (1989) hereby incorporated by reference. Synthetic biocompatible polymers are well known in the art and are also suitable implant materials. Examples of suitable synthetic polymers include urethanes, olefins, terephthalates, acrylates, polyesters and the like. Other acceptable implant materials are biodegradable hydrogels or aggregations of closely packed particles such as polymethylmethacrylate beads with a polymerized hydroxyethyl methacrylate coating. See the Encyclopedia of Polymer Science and Engineering, Vol. 2, pp. 267 et seq. (1989) hereby incorporated by reference.

The chimeric protein provides a useful way for immobilizing or coating a cellular regulatory factor on a pharmaceutically acceptable vehicle to deliver the cellular regulatory factor to desired sites in viable tissue. Suitable vehicles include those made of bioabsorbable polymers, biocompatible nonabsorbable polymers, lactoner putty and plaster of Paris. Examples of suitable bioabsorbable and biocompatible polymers include homopolymers, copolymers and blends of hydroxyacids such as lactide and glycolide, other absorbable polymers which may be used alone or in combination with hydroxyacids include dioxanones, carbonates such as trimethylene carbonate, lactones such as caprolactone, polyoxyaLkylenes, and oxylates. See the Encyclopedia of Polymer Science and Engineering, Vol. 2, pp. 230 et seq. (1989) hereby incorporated by reference.

These vehicles may be in the form of beads, particles, putty, coatings or film vehicles. Diffusional systems in which a core of chimeric protein is surrounded by a porous membrane layer are other acceptable vehicles.

The following examples should be considered as illustrative of certain embodiments disclosed herein but should not be considered as limiting the inventive disclosure.

EXAMPLE I

Cloning BMP-2B/collagen IA DNA segment constructs

Obtaining PCR products for BMP-2B and Collagen I(a): The chimeric gene encoding the BMP-2B/Collagen I(a) fusion protein is assembled from PCR products. The PCR primers are designed to provide restriction sites on the 5' and 3' ends that facilitate later ligation steps. The 5' and 3' ends of the BMP-2B PCR product contain BamHI and HindIII restriction sites respectively. The 5' and 3' ends of the Collagen I(a) PCR product contain XmnI and BglII restriction sites respectively. Amplification is carried out on template cDNA synthesized from total cellular RNA using standard methods. PCR reactions for BMP-2B and Collagen I(a) use cDNA prepared from U-20S and AG02261A cell lines respectively. After amplification and purification, the PCR products are ligated into PCR II vectors. Positive clones are identified by screening plasmids for the correct molecular weight. The clones are verified by DNA sequencing using standard methods. The BMP-2B PCR product is excised from PCRII by restriction digestion with BamHI and HindIII and the Collagen I(a) segment was excised from PCRII using XmnI and BglII. The restriction digest reactions are resolved by electrophoresis

through agarose gels and the DNA fragments with the BMP-2B and Collagen I(a) sequences are purified with gene clean (BIO 101).

EXAMPLE II

Cloning TGF-β_I/collagen IA DNA segment constructs

Obtaining PCR products for TGF-βI and Collagen I(a): The chimeric gene encoding the TGF-βI/Collagen I(a) fusion protein is assembled from PCR products. The PCR primers are designed to provide restriction sites on the 5' and 3' ends that facilitate later ligation steps. The 5' and 3' ends of the TGF-βI PCR product contain BglII and XbaI restriction sites respectively. The 5' and 3' ends of the Collagen I(a) PCR product contain XmnI and BglII restriction sites respectively. Amplification is carried out on template cDNA synthesized from total cellular RNA using standard methods. PCR reactions for TGF-β1 and Collagen I(a) use cDNA prepared from AG02261A cells. After amplification and purification, the PCR products are ligated into PCR II vectors. Positive clones are identified by screening plasmids for the correct molecular weight. The clones are verified by DNA sequencing using standard methods. The TGF-βI PCR product is excised from PCR II by restriction digestion with BglII and XbaI and the Collagen I(a) segment was excised from PCR II using XmnI and BglII. The restriction digest reactions are resolved by electrophoresis through agarose gels and the DNA fragments with the TGF-β1 and Collagen I(a) sequences are purified with gene clean (BIO 101).

EXAMPLE III

Cloning dermatan sulfate proteoglycan (decorin)/collagen IA DNA segment constructs

Obtaining PCR products for Decorin and Collagen I(a): The chimeric gene encoding the Decorin/Collagen I(a) fusion protein is assembled from PCR products. The PCR primers are designed to provide restriction sites on the 5' and 3' ends that facilitate later ligation steps. The 5' and 3' ends of the Decorin PCR product contain BamHI and PstI restriction sites respectively. The 5' and 3' ends of the Collagen I(a) PCR product contain XmnI and BglII restriction sites respectively. Amplification is carried out on template cDNA synthesized from total cellular RNA using standard methods. PCR reactions for Decorin and Collagen I(a) use cDNA prepared from AG02261A cells. After amplification and purification, the respective PCR products are ligated into respective PCR II vectors. Positive clones are identified by screening plasmids for the correct molecular weight. The clones are verified by DNA sequencing using standard methods. The Decorin PCR product is excised from PCR II by restriction digestion with BamHI and PstI and the Collagen I(a) segment was excised from PCR II using XmnI and BglII. The restriction digest reactions are resolved by electrophoresis through agarose gels and the DNA fragments with the Decorin and Collagen I(a) sequences are purified with gene clean (BIO101).

EXAMPLE IV

Construction of cloning vector incorporating DNA constructs of Example 1

Ligation of BMP-2B and Collagen I(a) segments into the pMal-c2 expression vector: The pMal-c2 vector is treated with BamHI and Hind3, resolved by electrophoresis through an agarose gel and purified by standard methods. The BMP-2B segment with matching BamHI and Hind3 restriction sites on the 5' and 3' ends is ligated into pMal-c2 and transformants are screened for the insert by standard techniques. Positive clones are verified by DNA sequencing and designated pMal-c2 BMP. To complete the construction, pMal-c2-BMP is digested with XmnI and BamHI and the Collagen I(a) segment which is digested with XmnI and BglII is ligated into those sites by standard methods (BamHI and BglII produce compatible termini). Positive clones are verified by DNA sequencing and designated pMal-CB. See Fig. 11.

EXAMPLE V

Construction of cloning vector incorporating DNA constructs of Example II

Ligation of TGF-B1 and Collagen I(a) segments into the pMal-c2 expression vector: The pMal-c2 vector is treated with XmnI and XbaI, resolved by electrophoresis through an agarose gel and purified by standard methods. The Collagen I(a) segment with a 5' XmnI site and a 3' BglII restriction site and the TGF-B1 segment with a 5' BglII site and a 3' XbaI site are combined with the digested and purified pMal-c2 plasmid for a three fragment ligation reaction using standard methods. Transformants are screened for the insert by standard techniques. Positive clones are verified by DNA sequencing and designated pMal-CT. See Fig. 12.

EXAMPLE VI

Construction of cloning vector incorporating DNA constructs of Example III

Ligation of Decorin and Collagen I(a) segments into the pMal-c2 expression vector: The pMal-c2 vector is treated with XmnI and PstI, resolved by electrophoresis through an agarose gel and purified by standard methods. The Collagen I(a) segment with a 5' XmnI site and a 3' BglII restriction site and the Decorin segment with a 5' BamHI site and a 3' PstI site are combined with the digested and purified pMal-c2 plasmid for a three fragment ligation reaction using standard methods (BamHI and BglII produce compatible termini). Transformants are screened for the insert by standard techniques. Positive clones are verified by DNA sequencing and designated pMal-CD. See Fig. 13.

EXAMPLE VII

Transformation of E. Coli and Expression of a Collagen/TGF-β 1 and Collagen/Decorin Chimeric Genes in E. coli

Expression plasmids pMal-CB (Collagen-BMP2B Chimera), pMal-CT (Collagen-TGF-BI Chimera) and pMal-CD (Collagen-Decorin Chimera) are used to transform E.coli HB 101 using standard techniques. To express protein, a 50 ml culture of E.coli harboring one of the expression vectors is inoculated into 1L of LB broth and incubated with agitation at 37°C. When the A_{600} is 0.5±0.1, 0.1M IPTG is added to a final concentration of 1.5-15 mM. The culture is maintained at 37°C until the A_{600} is 1.3 to 1.8 and the E.coli is harvested by centrifugation at 4000xg. The cell pellets are resuspended in 7.5 ml 20 mM Tris HCl pH 7.5, 200 mM NaCl, 1 mM EDTA (hereinafter "column buffer") and frozen in a dry ice/ethanol bath. The frozen cell pellets are thawed at 4°C, then sonicated on ice until the cells are disrupted. Cell debris is removed by centrifugation at 9,000xg at 4°C for 30 minutes. The supernatant fraction contains the E.coli crude cell lysate which is analyzed for protein production by SDS-PAGE. The recombinant protein products produced from these pMal vectors is a fusion protein with MBP (maltose binding protein). The MBP segment is included to allow a single step purification of the protein.

The crude lysate is passed over an amylose column containing ml of resin/3 mg of recombinant protein (expected yield). The column is washed with 8 volumes of column buffer and the column flow through is reapplied to the column. Another 8 volumes of column buffer is used to wash the column. The fusion protein is eluted from the column using column buffer containing 10 mM Maltose. Fractions containing the recombinant chimeric protein are identified by the BCA protein assay (Pierce) and verified by SDS-PAGE. The fractions that contain the protein are pooled

The MBP segment of the purified protein is cleaved from the collagen-growth factor chimera by treatment with factor Xa (New England Biolabs) at room temperature for 24 hours. The collagen-growth factor chimera is separated from the MBP segment by chromatography through an amylose column. The column flow through contains the collagen-growth factor chimera, which is analyzed by SDS-PAGE. Typical yield of purified protein range from 10-50 mg/liter of E.coli culture.

EXAMPLE VIII

Expression of a Collagen-Growth Factor Chimeric Genes in Sf9 Cells

A useful alternative to the E.coli expression system is Baculovirus. The gene for the collagen-growth factor chimeras is modified to include an ATG start codon at the 5' end and a TAA stop codon at the 3' end. The transcriptional unit is ligated into the baculoviral transfer vector pBlueBac III Invitrogen). The resulting transfer vector is verified by DNA sequencing. The collagen-growth factor chimera gene is transferred into the baculovirus genome (AcMNPV) by the standard in vivo recombination method. The pBlueBacIII transfer vector containing the collagen-growth factor chimera gene is cotransfected into Sf9 cells by standard methods. Recombinant viral plaques that are blue are selected and isolated by several rounds of reinfection. Pure recombinant baculovirus is verified by DNA sequencing. The recombinant virus containing the collagen-growth factor chimera gene is used to infect suspension cultures of Sf9 cells and optimal protein expression is determined at 48-72 hours post-infection. The protein product is recovered from the culture medium and analyzed by SDS-PAGE.

It will be understood that various modifications may be made to the embodiments disclosed herein. Therefore, the above description should not be construed as limiting, but merely as exemplifications of preferred embodiments. Those skilled in the art will envision other modifications within the scope and spirit of the claims appended hereto.

The claims which follow identify embodiments of the invention additional to those described in detail above.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: United States Surgical Corporation
(B) STREET: 150 Glover Avenue
(C) CITY: Norwalk
(D) STATE: Connecticut
(E) COUNTRY: USA
(F) POSTAL CODE (ZIP): 06856

(ii) TITLE OF INVENTION: Recombinant chimeric proteins and methods of use thereof

(iii) NUMBER OF SEQUENCES: 8

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(v) CURRENT APPLICATION DATA:
APPLICATION NUMBER: EP 95109019.0

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3535 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION 20..3526

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
GGGAAGGATT TCCATTTCCC AGCTGTCTTA TGGCTATGAT GAGAAATCAA CCGGAGGAAT      60
TTCCGTGCCT GGCCCCATGG GTCCCTCTGG TCCTCGTGGT CTCCCTGGCC CCCCTGGTGC     120
ACCTGGTCCC CAAGGCTTCC AAGGTCCCCC TGGTGAGCCT GGCGAGCCTG GAGCTTCAGG     180
TCCCATGGGT CCCCGAGGTC CCCCAGGTCC CCCTGGAAAG AATGGAGATG ATGGGGAAGC     240
TGGAAAACCT GGTCGTCCTG GTGAGCGTGG GCCTCCTGGG CCTCAGGGTG CTCGAGGATT     300
GCCCGGAACA GCTGGCCTCC CTGGAATGAA GGGACACAGA GGTTTCAGTG GTTTGGATGG     360
TGCCAAGGGA GATGCTGGTC CTGCTGGTCC TAAGGGTGAG CCTGGCAGCC CTGGTGAAAA     420
TGGAGCTCCT GGTCAGATGG GCCCCCGTGG CCTGCCTGGT GAGAGAGGTC GCCCTGGAGC     480
CCCTGGCCCT GCTGGTGCTC GTGGAAATGA TGGTGCTACT GGTGCTGCCG GGCCCCCTGG     540
TCCCACCGGC CCCGCTGGTC CTCCTGGCTT CCCTGGTGCT GTTGGTGCTA AGGGTGAAGC     600
```

```
TGGTCCCCAA GGGCCCCGAG GCTCTGAAGG TCCCCAGGGT GTGCGTGGTG AGCCTGGCCC    660
CCCTGGCCCT GCTGGTGCTG CTGGCCCTGC TGGAAACCCT GGTGCTGATG GACAGCCTGG    720
TGCTAAAGGT GCCAATGGTG CTCCTGGTAT TGCTGGTGCT CCTGGCTTCC CTGGTGCCCG    780
AGGCCCCTCT GGACCCCAGG GCCCCGGCGG CCCTCCTGGT CCCAAGGGTA ACAGCGGTGA    840
ACCTGGTGCT CCTGGCAGCA AAGGAGACAC TGGTGCTAAG GGAGAGCCTG GCCCTGTTGG    900
TGTTCAAGGA CCCCCTGGCC CTGCTGGAGA GGAAGGAAAG CGAGGAGCTC GAGGTGAACC    960
CGGACCCACT GGCCTGCCCG GACCCCCTGG CGAGCGTGGT GGACCTGGTA GCCGTGGTTT   1020
CCCTGGCGCA GATGGTGTTG CTGGTCCCAA GGGTCCCGCT GGTGAACGTG GTTCTCCTGG   1080
CCCCGCTGGC CCCAAAGGAT CTCCTGGTGA AGCTGGTCGT CCCGGTGAAG CTGGTCTGCC   1140
TGGTGCCAAG GGTCTGACTG GAAGCCCTGG CAGCCCTGGT CCTGATGGCA AAACTGGCCC   1200
CCCTGGTCCC GCCGGTCAAG ATGGTCGCCC CGGACCCCCA GGCCCACCTG GTGCCCGTGG   1260
TCAGGCTGGT GTGATGGGAT TCCCTGGACC TAAAGGTGCT GCTGGAGAGC CCGGCAAGGC   1320
TGGAGAGCGA GGTGTTCCCG GACCCCCTGG CGCTGTCGGT CCTGCTGGCA AAGATGGAGA   1380
GGCTGGAGCT CAGGGACCCC CTGGCCCTGC TGGTCCCGCT GGCGAGAGAG GTGAACAAGG   1440
CCCTGCTGGC TCCCCCGGAT TCCAGGGTCT CCCTGGTCCT GCTGGTCCTC CAGGTGAAGC   1500
AGGCAAACCT GGTGAACAGG GTGTTCCTGG AGACCTTGGC GCCCCTGGCC CCTCTGGAGC   1560
AAGAGGCGAG AGAGGTTTCC CTGGCGAGCG TGGTGTGCAA GGTCCCCCTG GTCCTGCTGG   1620
ACCCCGAGGG GCCAACGGTG CTCCCGGCAA CGATGGTGCT AAGGGTGATG CTGGTGCCCC   1680
TGGAGCTCCC GGTAGCCAGG GCGCCCCTGG CCTTCAGGGA ATGCCTGGTG AACGTGGTGC   1740
AGCTGGTCTT CCAGGGCCTA AGGGTGACAG AGGTGATGCT GGTCCCAAAG GTGCTGATGG   1800
CTCTCCTGGC AAAGATGGCG TCCGTGGTCT GACCGGCCCC ATTGGTCCTC CTGGCCCTGC   1860
TGGTGCCCCT GGTGACAAGG GTGAAAGTGG TCCCAGCGGC CCTGCTGGTC CCACTGGAGC   1920
TCGTGGTGCC CCCGGAGACC GTGGTGAGCC TGGTCCCCCC GGCCCTGCTG GCTTTGCTGG   1980
CCCCCCTGGT GCTGACGGCC AACCTGGTGC TAAAGGCGAA CCTGGTGATG CTGGTGCCAA   2040
AGGCGATGCT GGTCCCCCTG GCCCTGCCGG ACCCGCTGGA CCCCCTGGCC CCATTGGTAA   2100
TGTTGGTGCT CCTGGAGCCA AAGGTGCTCG CGGCAGCGCT GGTCCCCCTG GTGCTACTGG   2160
TTTCCCTGGT GCTGCTGGCC GAGTCGGTCC TCCTGGCCCC TCTGGAAATG CTGGACCCCC   2220
TGGCCCTCCT GGTCCTGCTG GCAAAGAAGG CGGCAAAGGT CCCCGTGGTG AGACTGGCCC   2280
TGCTGGACGT CCTGGTGAAG TTGGTCCCCC TGGTCCCCCT GGCCCTGCTG GCGAGAAAGG   2340
ATCCCCTGGT GCTGATGGTC CTGCTGGTGC TCCTGGTACT CCCGGGCCTC AAGGTATTGC   2400
TGGACAGCGT GGTGTGGTCG GCCTGCCTGG TCAGAGAGGA GAGAGAGGCT TCCCTGGTCT   2460
TCCTGGCCCC TCTGGTGAAC CTGGCAAACA AGGTCCCTCT GGAGCAAGTG GTGAACGTGG   2520
TCCCCCCGGT CCCATGGGCC CCCCTGGATT GGCTGGACCC CCTGGTGAAT CTGGACGTGA   2580
```

GGGGGCTCCT GCTGCCGAAG GTTCCCCTGG ACGAGACGGT TCTCCTGGCG CCAAGGGTGA 2640

CCGTGGTGAG ACCGGCCCCG CTGGACCCCC TGGTGCTCNT GGTGCTCNTG GTGCCCCTGG 2700

CCCCGTTGGC CCTGCTGGCA AGAGTGGTGA TCGTGGTGAG ACTGGTCCTG CTGGTCCCGC 2760

CGGTCCCGTC GGCCCCGCTG GCGCCCGTGG CCCCGCCGGA CCCCAAGGCC CCCGTGGTGA 2820

CAAGGGTGAG ACAGGCGAAC AGGGCGACAG AGGCATAAAG GGTCACCGTG GCTTCTCTGG 2880

CCTCCAGGGT CCCCCTGGCC CTCCTGGCTC TCCTGGTGAA CAAGGTCCCT CTGGAGCCTC 2940

TGGTCCTGCT GGTCCCCGAG GTCCCCCTGG CTCTGCTGGT GCTCCTGGCA AAGATGGACT 3000

CAACGGTCTC CCTGGCCCCA TTGGGCCCCC TGGTCCTCGC GGTCGCACTG GTGATGCTGG 3060

TCCTGTTGGT CCCCCCGGCC CTCCTGGACC TCCTGGTCCC CCTGGTCCTC CCAGCGCTGG 3120

TTTCGACTTC AGCTTCCTCC CCCAGCCACC TCAAGAGAAG GCTCACGATG GTGGCCGCTA 3180

CTACCGGGCT AGATCCCAGC GGGCCAGGAA GAAGAATAAG AACTGCCGGC GCCACTCGCT 3240

CTATGTGGAC TTCAGCGATG TGGGCTGGAA TGACTGGATT GTGGCCCCAC CAGGCTACCA 3300

GGCCTTCTAC TGCCATGGGG ACTGCCCCTT TCCACTGGCT GACCACCTCA ACTCAACCAA 3360

CCATGCCATT GTGCAGACCC TGGTCAATTC TGTCAATTCC AGTATCCCCA AAGCCTGTTG 3420

TGTGCCCACT GAACTGAGTG CCATCTCCAT GCTGTACCTG GATGAGTATG ATAAGGTGGT 3480

ACTGAAAAAT TATCAGGAGA TGGTAGTAGA GGGATGTGGG TGCCGCTAAA AGCTT 3535

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3541 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:20..3532

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

GGGAAGGATT TCCATTTCCC AGCTGTCTTA TGGCTATGAT GAGAAATCAA CCGGAGGAAT 60

TTCCGTGCCT GGCCCCATGG GTCCCTCTGG TCCTCGTGGT CTCCCTGGCC CCCCTGGTGC 120

ACCTGGTCCC CAAGGCTTCC AAGGTCCCCC TGGTGAGCCT GGCGAGCCTG GAGCTTCAGG 180

TCCCATGGGT CCCCGAGGTC CCCCAGGTCC CCCTGGAAAG AATGGAGATG ATGGGGAAGC 240

TGGAAAACCT GGTCGTCCTG GTGAGCGTGG GCCTCCTGGG CCTCAGGGTG CTCGAGGATT 300

GCCCGGAACA GCTGGCCTCC CTGGAATGAA GGGACACAGA GGTTTCAGTG GTTTGGATGG 360

TGCCAAGGGA GATGCTGGTC CTGCTGGTCC TAAGGGTGAG CCTGGCAGCC CTGGTGAAAA 420
TGGAGCTCCT GGTCAGATGG GCCCCCGTGG CCTGCCTGGT GAGAGAGGTC GCCCTGGAGC 480
CCCTGGCCCT GCTGGTGCTC GTGGAAATGA TGGTGCTACT GGTGCTGCCG GGCCCCCTGG 540
TCCCACCGGC CCCGCTGGTC CTCCTGGCTT CCCTGGTGCT GTTGGTGCTA AGGGTGAAGC 600
TGGTCCCCAA GGGCCCCGAG GCTCTGAAGG TCCCCAGGGT GTGCGTGGTG AGCCTGGCCC 660
CCCTGGCCCT GCTGGTGCTG CTGGCCCTGC TGGAAACCCT GGTGCTGATG GACAGCCTGG 720
TGCTAAAGGT GCCAATGGTG CTCCTGGTAT TGCTGGTGCT CCTGGCTTCC CTGGTGCCCG 780
AGGCCCCTCT GGACCCCAGG GCCCCGGCGG CCCTCCTGGT CCCAAGGGTA ACAGCGGTGA 840
ACCTGGTGCT CCTGGCAGCA AAGGAGACAC TGGTGCTAAG GGAGAGCCTG GCCCTGTTGG 900
TGTTCAAGGA CCCCCTGGCC CTGCTGGAGA GGAAGGAAAG CGAGGAGCTC GAGGTGAACC 960
CGGACCCACT GGCCTGCCCG GACCCCCTGG CGAGCGTGGT GGACCTGGTA GCCGTGGTTT 1020
CCCTGGCGCA GATGGTGTTG CTGGTCCCAA GGGTCCCGCT GGTGAACGTG GTTCTCCTGG 1080
CCCCGCTGGC CCCAAAGGAT CTCCTGGTGA AGCTGGTCGT CCCGGTGAAG CTGGTCTGCC 1140
TGGTGCCAAG GGTCTGACTG GAAGCCCTGG CAGCCCTGGT CCTGATGGCA AAACTGGCCC 1200
CCCTGGTCCC GCCGGTCAAG ATGGTCGCCC CGGACCCCCA GGCCCACCTG GTGCCCGTGG 1260
TCAGGCTGGT GTGATGGGAT TCCCTGGACC TAAAGGTGCT GCTGGAGAGC CCGGCAAGGC 1320
TGGAGAGCGA GGTGTTCCCG GACCCCCTGG CGCTGTCGGT CCTGCTGGCA AAGATGGAGA 1380
GGCTGGAGCT CAGGGACCCC CTGGCCCTGC TGGTCCCGCT GGCGAGAGAG GTGAACAAGG 1440
CCCTGCTGGC TCCCCCGGAT TCCAGGGTCT CCCTGGTCCT GCTGGTCCTC CAGGTGAAGC 1500
AGGCAAACCT GGTGAACAGG GTGTTCCTGG AGACCTTGGC GCCCCTGGCC CCTCTGGAGC 1560
AAGAGGCGAG AGAGGTTTCC CTGGCGAGCG TGGTGTGCAA GGTCCCCCTG GTCCTGCTGG 1620
ACCCCGAGGG GCCAACGGTG CTCCCGGCAA CGATGGTGCT AAGGGTGATG CTGGTGCCCC 1680
TGGAGCTCCC GGTAGCCAGG GCGCCCCTGG CCTTCAGGGA ATGCCTGGTG AACGTGGTGC 1740
AGCTGGTCTT CCAGGGCCTA AGGGTGACAG AGGTGATGCT GGTCCCAAAG GTGCTGATGG 1800
CTCTCCTGGC AAAGATGGCG TCCGTGGTCT GACCGGCCCC ATTGGTCCTC CTGGCCCTGC 1860
TGGTGCCCCT GGTGACAAGG GTGAAAGTGG TCCCAGCGGC CCTGCTGGTC CCACTGGAGC 1920
TCGTGGTGCC CCCGGAGACC GTGGTGAGCC TGGTCCCCCC GGCCCTGCTG GCTTTGCTGG 1980
CCCCCCTGGT GCTGACGGCC AACCTGGTGC TAAAGGCGAA CCTGGTGATG CTGGTGCCAA 2040
AGGCGATGCT GGTCCCCCTG GGCCTGCCGG ACCCGCTGGA CCCCCTGGCC CCATTGGTAA 2100
TGTTGGTGCT CCTGGAGCCA AAGGTGCTCG CGGCAGCGCT GGTCCCCCTG GTGCTACTGG 2160
TTTCCCTGGT GCTGCTGGCC GAGTCGGTCC TCCTGGCCCC TCTGGAAATG CTGGACCCCC 2220
TGGCCCTCCT GGTCCTGCTG GCAAAGAAGG CGGCAAAGGT CCCCGTGGTG AGACTGGCCC 2280
TGCTGGACGT CCTGGTGAAG TTGGTCCCCC TGGTCCCCCT GGCCCTGCTG GCGAGAAAGG 2340

```
ATCCCCTGGT GCTGATGGTC CTGCTGGTGC TCCTGGTACT CCCGGGCCTC AAGGTATTGC    2400
TGGACAGCGT GGTGTGGTCG GCCTGCCTGG TCAGAGAGGA GAGAGAGGCT TCCCTGGTCT    2460
TCCTGGCCCC TCTGGTGAAC CTGGCAAACA AGGTCCCTCT GGAGCAAGTG GTGAACGTGG    2520
TCCCCCCGGT CCCATGGGCC CCCCTGGATT GGCTGGACCC CCTGGTGAAT CTGGACGTGA    2580
GGGGGCTCCT GCTGCCGAAG GTTCCCCTGG ACGAGACGGT TCTCCTGGCG CCAAGGGTGA    2640
CCGTGGTGAG ACCGGCCCCG CTGGACCCCC TGGTGCTCNT GGTGCTCNTG GTGCCCCTGG    2700
CCCCGTTGGC CCTGCTGGCA AGAGTGGTGA TCGTGGTGAG ACTGGTCCTG CTGGTCCCGC    2760
CGGTCCCGTC GGCCCCGCTG GCGCCCGTGG CCCCGCCGGA CCCCAAGGCC CCCGTGGTGA    2820
CAAGGGTGAG ACAGGCGAAC AGGGCGACAG AGGCATAAAG GGTCACCGTG GCTTCTCTGG    2880
CCTCCAGGGT CCCCCTGGCC CTCCTGGCTC TCCTGGTGAA CAAGGTCCCT CTGGAGCCTC    2940
TGGTCCTGCT GGTCCCCGAG GTCCCCCTGG CTCTGCTGGT GCTCCTGGCA AAGATGGACT    3000
CAACGGTCTC CCTGGCCCCA TTGGGCCCCC TGGTCCTCGC GGTCGCACTG GTGATGCTGG    3060
TCCTGTTGGT CCCCCCGGCC CTCCTGGACC TCCTGGTCCC CCTGGTCCTC CCAGCGCTGG    3120
TTTCGACTTC AGCTTCCTCC CCCAGCCACC TCAAGAGAAG GCTCACGATG GTGGCCGCTA    3180
CTACCGGGCT AGATCTGCCC TGGACACCAA CTATTGCTTC AGCTCCACGG AGAAGAACTG    3240
CTGCGTGCGG CAGCTGTACA TTGACTTCCG CAAGGACCTC GGCTGGAAGT GGATCCACGA    3300
GCCCAAGGGC TACCATGCCA ACTTCTGCCT CGGGCCCTGC CCCTACATTT GGAGCCTGGA    3360
CACGCAGTAC AGCAAGGTCC TGGCCCTGTA CAACCAGCAT AACCCGGGCG CCTCGGCGGC    3420
GCCGTGCTGC GTGCCGCAGG CGCTGGAGCC GCTGCCCATC GTGTACTACG TGGGCCGCAA    3480
GCCCAAGGTG GAGCAGCTGT CCAACATGAT CGTGCGCTCC TGCAAGTGCA GCTGATCTAG    3540
A                                                                    3541
```

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 4192 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
- (A) ORGANISM: Homo sapiens

(ix) FEATURE:
- (A) NAME/KEY: CDS
- (B) LOCATION:20..4183

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```
GGGAAGGATT TCCATTTCCC AGCTGTCTTA TGGCTATGAT GAGAAATCAA CCGGAGGAAT      60
TTCCGTGCCT GGCCCCATGG GTCCCTCTGG TCCTCGTGGT CTCCCTGGCC CCCCTGGTGC     120
```

```
ACCTGGTCCC CAAGGCTTCC AAGGTCCCCC TGGTGAGCCT GGCGAGCCTG GAGCTTCAGG    180
TCCCATGGGT CCCCGAGGTC CCCCAGGTCC CCCTGGAAAG AATGGAGATG ATGGGGAAGC    240
TGGAAAACCT GGTCGTCCTG GTGAGCGTGG GCCTCCTGGG CCTCAGGGTG CTCGAGGATT    300
GCCCGGAACA GCTGGCCTCC CTGGAATGAA GGGACACAGA GGTTTCAGTG GTTTGGATGG    360
TGCCAAGGGA GATGCTGGTC CTGCTGGTCC TAAGGGTGAG CCTGGCAGCC CTGGTGAAAA    420
TGGAGCTCCT GGTCAGATGG GCCCCCGTGG CCTGCCTGGT GAGAGAGGTC GCCCTGGAGC    480
CCCTGGCCCT GCTGGTGCTC GTGGAAATGA TGGTGCTACT GGTGCTGCCG GGCCCCCTGG    540
TCCCACCGGC CCCGCTGGTC CTCCTGGCTT CCCTGGTGCT GTTGGTGCTA AGGGTGAAGC    600
TGGTCCCCAA GGGCCCCGAG GCTCTGAAGG TCCCCAGGGT GTGCGTGGTG AGCCTGGCCC    660
CCCTGGCCCT GCTGGTGCTG CTGGCCCTGC TGGAAACCCT GGTGCTGATG GACAGCCTGG    720
TGCTAAAGGT GCCAATGGTG CTCCTGGTAT TGCTGGTGCT CCTGGCTTCC CTGGTGCCCG    780
AGGCCCCTCT GGACCCCAGG GCCCCGGCGG CCCTCCTGGT CCCAAGGGTA ACAGCGGTGA    840
ACCTGGTGCT CCTGGCAGCA AAGGAGACAC TGGTGCTAAG GGAGAGCCTG GCCCTGTTGG    900
TGTTCAAGGA CCCCCTGGCC CTGCTGGAGA GGAAGGAAAG CGAGGAGCTC GAGGTGAACC    960
CGGACCCACT GGCCTGCCCG GACCCCCTGG CGAGCGTGGT GGACCTGGTA GCCGTGGTTT   1020
CCCTGGCGCA GATGGTGTTG CTGGTCCCAA GGGTCCCGCT GGTGAACGTG GTTCTCCTGG   1080
CCCCGCTGGC CCCAAAGGAT CTCCTGGTGA AGCTGGTCGT CCCGGTGAAG CTGGTCTGCC   1140
TGGTGCCAAG GGTCTGACTG GAAGCCCTGG CAGCCCTGGT CCTGATGGCA AAACTGGCCC   1200
CCCTGGTCCC GCCGGTCAAG ATGGTCGCCC CGGACCCCCA GGCCCACCTG GTGCCCGTGG   1260
TCAGGCTGGT GTGATGGGAT TCCCTGGACC TAAAGGTGCT GCTGGAGAGC CCGGCAAGGC   1320
TGGAGAGCGA GGTGTTCCCG GACCCCCTGG CGCTGTCGGT CCTGCTGGCA AAGATGGAGA   1380
GGCTGGAGCT CAGGGACCCC CTGGCCCTGC TGGTCCCGCT GGCGAGAGAG GTGAACAAGG   1440
CCCTGCTGGC TCCCCCGGAT TCCAGGGTCT CCCTGGTCCT GCTGGTCCTC CAGGTGAAGC   1500
AGGCAAACCT GGTGAACAGG GTGTTCCTGG AGACCTTGGC GCCCCTGGCC CCTCTGGAGC   1560
AAGAGGCGAG AGAGGTTTCC CTGGCGAGCG TGGTGTGCAA GGTCCCCCTG GTCCTGCTGG   1620
ACCCCGAGGG GCCAACGGTG CTCCCGGCAA CGATGGTGCT AAGGGTGATG CTGGTGCCCC   1680
TGGAGCTCCC GGTAGCCAGG GCGCCCCTGG CCTTCAGGGA ATGCCTGGTG AACGTGGTGC   1740
AGCTGGTCTT CCAGGGCCTA AGGGTGACAG AGGTGATGCT GGTCCCAAAG GTGCTGATGG   1800
CTCTCCTGGC AAAGATGGCG TCCGTGGTCT GACCGGCCCC ATTGGTCCTC CTGGCCCTGC   1860
TGGTGCCCCT GGTGACAAGG GTGAAAGTGG TCCCAGCGGC CCTGCTGGTC CCACTGGAGC   1920
TCGTGGTGCC CCCGGAGACC GTGGTGAGCC TGGTCCCCCC GGCCCTGCTG GCTTTGCTGG   1980
CCCCCCTGGT GCTGACGGCC AACCTGGTGC TAAAGGCGAA CCTGGTGATG CTGGTGCCAA   2040
AGGCGATGCT GGTCCCCCTG GGCCTGCCGG ACCCGCTGGA CCCCCTGGCC CCATTGGTAA   2100
```

TGTTGGTGCT CCTGGAGCCA AAGGTGCTCG CGGCAGCGCT GGTCCCCCTG GTGCTACTGG 2160
TTTCCCTGGT GCTGCTGGCC GAGTCGGTCC TCCTGGCCCC TCTGGAAATG CTGGACCCCC 2220
TGGCCCTCCT GGTCCTGCTG GCAAAGAAGG CGGCAAAGGT CCCCGTGGTG AGACTGGCCC 2280
TGCTGGACGT CCTGGTGAAG TTGGTCCCCC TGGTCCCCCT GGCCCTGCTG GCGAGAAAGG 2340
ATCCCCTGGT GCTGATGGTC CTGCTGGTGC TCCTGGTACT CCCGGGCCTC AAGGTATTGC 2400
TGGACAGCGT GGTGTGGTCG GCCTGCCTGG TCAGAGAGGA GAGAGAGGCT TCCCTGGTCT 2460
TCCTGGCCCC TCTGGTGAAC CTGGCAAACA AGGTCCCTCT GGAGCAAGTG GTGAACGTGG 2520
TCCCCCCGGT CCCATGGGCC CCCCTGGATT GGCTGGACCC CCTGGTGAAT CTGGACGTGA 2580
GGGGGCTCCT GCTGCCGAAG GTTCCCCTGG ACGAGACGGT TCTCCTGGCG CCAAGGGTGA 2640
CCGTGGTGAG ACCGGCCCCG CTGGACCCCC TGGTGCTCNT GGTGCTCNTG GTGCCCCTGG 2700
CCCCGTTGGC CCTGCTGGCA AGAGTGGTGA TCGTGGTGAG ACTGGTCCTG CTGGTCCCGC 2760
CGGTCCCGTC GGCCCCGCTG GCGCCCGTGG CCCCGCCGGA CCCCAAGGCC CCCGTGGTGA 2820
CAAGGGTGAG ACAGGCGAAC AGGGCGACAG AGGCATAAAG GGTCACCGTG GCTTCTCTGG 2880
CCTCCAGGGT CCCCCTGGCC CTCCTGGCTC TCCTGGTGAA CAAGGTCCCT CTGGAGCCTC 2940
TGGTCCTGCT GGTCCCCGAG GTCCCCCTGG CTCTGCTGGT GCTCCTGGCA AAGATGGACT 3000
CAACGGTCTC CCTGGCCCCA TTGGGCCCCC TGGTCCTCGC GGTCGCACTG GTGATGCTGG 3060
TCCTGTTGGT CCCCCCGGCC CTCCTGGACC TCCTGGTCCC CCTGGTCCTC CCAGCGCTGG 3120
TTTCGACTTC AGCTTCCTCC CCCAGCCACC TCAAGAGAAG GCTCACGATG GTGGCCGCTA 3180
CTACCGGGCT AGATCTGATG AGGCTTCTGG GATAGGCCCA GAAGTTCCTG ATGACCGCGA 3240
CTTCGAGCCC TCCCTAGGCC CAGTGTGCCC CTTCCGCTGT CAATGCCATC TTCGAGTGGT 3300
CCAGTGTTCT GATTTGGGTC TGGACAAAGT GCCAAAGGAT CTTCCCCCTG ACACAACTCT 3360
GCTAGACCTG CAAAACAACA AAATAACCGA AATCAAAGAT GGAGACTTTA AGAACCTGAA 3420
GAACCTTCAC GCATTGATTC TTGTCAACAA TAAAATTAGC AAAGTTAGTC CTGGAGCATT 3480
TACACCTTTG GTGAAGTTGG AACGACTTTA TCTGTCCAAG AATCAGCTGA AGGAATTGCC 3540
AGAAAAAATG CCCAAAACTC TTCAGGAGCT GCGTGCCCAT GAGAATGAGA TCACCAAAGT 3600
GCGAAAAGTT ACTTTCAATG GACTGAACCA GATGATTGTC ATAGAACTGG GCACCAATCC 3660
GCTGAAGAGC TCAGGAATTG AAAATGGGGC TTTCCAGGGA ATGAAGAAGC TCTCCTACAT 3720
CCGCATTGCT GATACCAATA TCACCAGCAT TCCTCAAGGT CTTCCTCCTT CCCTTACGGA 3780
ATTACATCTT GATGGCAACA AAATCAGCAG AGTTGATGCA GCTAGCCTGA AAGGACTGAA 3840
TAATTTGGCT AAGTTGGGAT TGAGTTTCAA CAGCATCTCT GCTGTTGACA ATGGCTCTCT 3900
GGCCAACACG CCTCATCTGA GGGAGCTTCA CTTGGACAAC AACAAGCTTA CCAGAGTACC 3960
TGGTGGGCTG GCAGAGCATA AGTACATCCA GGTTGTCTAC CTTCATAACA ACAATATCTC 4020
TGTAGTTGGA TCAAGTGACT TCTGCCCACC TGGACACAAC ACCAAAAAGG CTTCTTATTC 4080

GGGTGTGAGT CTTTTCAGCA ACCCGGTCCA GTACTGGGAG ATACAGCCAT CCACCTTCAG 4140

ATGTGTCTAC GTGCGCTCTG CCATTCAACT CGGAAACTAT AAGTAACTGC AG 4192

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3349 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:20..3340

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

GGGAAGGATT TCCATTTCCC AGCTGTCTTA TGGCTATGAT GAGAAATCAA CCGGAGGAAT 60

TTCCGTGCCT GGCCCCATGG GTCCCTCTGG TCCTCGTGGT CTCCCTGGCC CCCCTGGTGC 120

ACCTGGTCCC CAAGGCTTCC AAGGTCCCCC TGGTGAGCCT GGCGAGCCTG GAGCTTCAGG 180

TCCCATGGGT CCCCGAGGTC CCCCAGGTCC CCCTGGAAAG AATGGAGATG ATGGGGAAGC 240

TGGAAAACCT GGTCGTCCTG GTGAGCGTGG GCCTCCTGGG CCTCAGGGTG CTCGAGGATT 300

GCCCGGAACA GCTGGCCTCC CTGGAATGAA GGGACACAGA GGTTTCAGTG GTTTGGATGG 360

TGCCAAGGGA GATGCTGGTC CTGCTGGTCC TAAGGGTGAG CCTGGCAGCC CTGGTGAAAA 420

TGGAGCTCCT GGTCAGATGG GCCCCCGTGG CCTGCCTGGT GAGAGAGGTC GCCCTGGAGC 480

CCCTGGCCCT GCTGGTGCTC GTGGAAATGA TGGTGCTACT GGTGCTGCCG GGCCCCCTGG 540

TCCCACCGGC CCCGCTGGTC CTCCTGGCTT CCCTGGTGCT GTTGGTGCTA AGGGTGAAGC 600

TGGTCCCCAA GGGCCCCGAG GCTCTGAAGG TCCCCAGGGT GTGCGTGGTG AGCCTGGCCC 660

CCCTGGCCCT GCTGGTGCTG CTGGCCCTGC TGGAAACCCT GGTGCTGATG GACAGCCTGG 720

TGCTAAAGGT GCCAATGGTG CTCCTGGTAT TGCTGGTGCT CCTGGCTTCC CTGGTGCCCG 780

AGGCCCCTCT GGACCCCAGG GCCCCGGCGG CCCTCCTGGT CCCAAGGGTA ACAGCGGTGA 840

ACCTGGTGCT CCTGGCAGCA AAGGAGACAC TGGTGCTAAG GGAGAGCCTG GCCCTGTTGG 900

TGTTCAAGGA CCCCCTGGCC CTGCTGGAGA GGAAGGAAAG CGAGGAGCTC GAGGTGAACC 960

CGGACCCACT GGCCTGCCCG GACCCCCTGG CGAGCGTGGT GGACCTGGTA GCCGTGGTTT 1020

CCCTGGCGCA GATGGTGTTG CTGGTCCCAA GGGTCCCGCT GGTGAACGTG GTTCTCCTGG 1080

CCCCGCTGGC CCCAAAGGAT CTCCTGGTGA AGCTGGTCGT CCCGGTGAAG CTGGTCTGCC 1140

TGGTGCCAAG GGTCTGACTG GAAGCCCTGG CAGCCCTGGT CCTGATGGCA AAACTGGCCC 1200

CCCTGGTCCC GCCGGTCAAG ATGGTCGCCC CGGACCCCCA GGCCCACCTG GTGCCCGTGG 1260

TCAGGCTGGT GTGATGGGAT TCCCTGGACC TAAAGGTGCT GCTGGAGAGC CCGGCAAGGC 1320

```
TGGAGAGCGA GGTGTTCCCG GACCCCCTGG CGCTGTCGGT CCTGCTGGCA AAGATGGAGA   1380
GGCTGGAGCT CAGGGACCCC CTGGCCCTGC TGGTCCCGCT GGCGAGAGAG GTGAACAAGG   1440
CCCTGCTGGC TCCCCCGGAT TCCAGGGTCT CCCTGGTCCT GCTGGTCCTC CAGGTGAAGC   1500
AGGCAAACCT GGTGAACAGG GTGTTCCTGG AGACCTTGGC GCCCCTGGCC CCTCTGGAGC   1560
AAGAGGCGAG AGAGGTTTCC CTGGCGAGCG TGGTGTGCAA GGTCCCCCTG GTCCTGCTGG   1620
ACCCCGAGGG GCCAACGGTG CTCCCGGCAA CGATGGTGCT AAGGGTGATG CTGGTGCCCC   1680
TGGAGCTCCC GGTAGCCAGG GCGCCCCTGG CCTTCAGGGA ATGCCTGGTG AACGTGGTGC   1740
AGCTGGTCTT CCAGGGCCTA AGGGTGACAG AGGTGATGCT GGTCCCAAAG GTGCTGATGG   1800
CTCTCCTGGC AAAGATGGCG TCCGTGGTCT GACCGGCCCC ATTGGTCCTC CTGGCCCTGC   1860
TGGTGCCCCT GGTGACAAGG GTGAAAGTGG TCCCAGCGGC CCTGCTGGTC CCACTGGAGC   1920
TCGTGGTGCC CCCGGAGACC GTGGTGAGCC TGGTCCCCCC GGCCCTGCTG GCTTTGCTGG   1980
CCCCCCTGGT GCTGACGGCC AACCTGGTGC TAAAGGCGAA CCTGGTGATG CTGGTGCCAA   2040
AGGCGATGCT GGTCCCCCTG GCCCTGCCGG ACCCGCTGGA CCCCCTGGCC CCATTGGTAA   2100
TGTTGGTGCT CCTGGAGCCA AAGGTGCTCG CGGCAGCGCT GGTCCCCCTG GTGCTACTGG   2160
TTTCCCTGGT GCTGCTGGCC GAGTCGGTCC TCCTGGCCCC TCTGGAAATG CTGGACCCCC   2220
TGGCCCTCCT GGTCCTGCTG GCAAAGAAGG CGGCAAAGGT CCCCGTGGTG AGACTGGCCC   2280
TGCTGGACGT CCTGGTGAAG TTGGTCCCCC TGGTCCCCCT GGCCCTGCTG GCGAGAAAGG   2340
ATCCCCTGGT GCTGATGGTC CTGCTGGTGC TCCTGGTACT CCCGGGCCTC AAGGTATTGC   2400
TGGACAGCGT GGTGTGGTCG GCCTGCCTGG TCAGAGAGGA GAGAGAGGCT TCCCTGGTCT   2460
TCCTGGCCCC TCTGGTGAAC CTGGCAAACA AGGTCCCTCT GGAGCAAGTG GTGAACGTGG   2520
TCCCCCCGGT CCCATGGGCC CCCCTGGATT GGCTGGACCC CCTGGTGAAT CTGGACGTGA   2580
GGGGGCTCCT GCTGCCGAAG GTTCCCCTGG ACGAGACGGT TCTCCTGGCG CCAAGGGTGA   2640
CCGTGGTGAG ACCGGCCCCG CTGGACCCCC TGGTGCTCNT GGTGCTCNTG GTGCCCCTGG   2700
CCCCGTTGGC CCTGCTGGCA AGAGTGGTGA TCGTGGTGAG ACTGGTCCTG CTGGTCCCGC   2760
CGGTCCCGTC GGCCCCGCTG GCGCCCGTGG CCCCGCCGGA CCCCAAGGCC CCCGTGGTGA   2820
CAAGGGTGAG ACAGGCGAAC AGGGCGACAG AGGCATAAAG GGTCACCGTG GCTTCTCTGG   2880
CCTCCAGGGT CCCCCTGGCC CTCCTGGCTC TCCTGGTGAA CAAGGTCCCT CTGGAGCCTC   2940
TGGTCCTGCT GGTCCCCGAG GTCCCCCTGG CTCTGCTGGT GCTCCTGGCA AAGATGGACT   3000
CAACGGTCTC CCTGGCCCCA TTGGGCCCCC TGGTCCTCGC GGTCGCACTG GTGATGCTGG   3060
TCCTGTTGGT CCCCCCGGCC CTCCTGGACC TCCTGGTCCC CCTGGTCCTC CCAGCGCTGG   3120
TTTCGACTTC AGCTTCCTCC CCCAGCCACC TCAAGAGAAG GCTCACGATG GTGGCCGCTA   3180
CTACCGGGCT AGATCTCCAA AGGATCTTCC CCCTGACACA ACTCTGCTAG ACCTGCAAAA   3240
CAACAAAATA ACCGAAATCA AAGATGGAGA CTTTAAGAAC CTGAAGAACC TTCACGCATT   3300
```

GATTCTTGTC AACAATAAAA TTAGCAAAGT TAGTCCTGGA TAACTGCAG 3349

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1169 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

```
Gln Leu Ser Tyr Gly Tyr Asp Glu Lys Ser Thr Gly Gly Ile Ser Val
  1               5                  10                  15

Pro Gly Pro Met Gly Pro Ser Gly Pro Arg Gly Leu Pro Gly Pro Pro
             20                  25                  30

Gly Ala Pro Gly Pro Gln Gly Phe Gln Gly Pro Pro Gly Glu Pro Gly
         35                  40                  45

Glu Pro Gly Ala Ser Gly Pro Met Gly Pro Arg Gly Pro Pro Gly Pro
     50                  55                  60

Pro Gly Lys Asn Gly Asp Asp Gly Glu Ala Gly Lys Pro Gly Arg Pro
 65                  70                  75                  80

Gly Glu Arg Gly Pro Pro Gly Pro Gln Gly Ala Arg Gly Leu Pro Gly
                 85                  90                  95

Thr Ala Gly Leu Pro Gly Met Lys Gly His Arg Gly Phe Ser Gly Leu
            100                 105                 110

Asp Gly Ala Lys Gly Asp Ala Gly Pro Ala Gly Pro Lys Gly Glu Pro
        115                 120                 125

Gly Ser Pro Gly Glu Asn Gly Ala Pro Gly Gln Met Gly Pro Arg Gly
    130                 135                 140

Leu Pro Gly Glu Arg Gly Arg Pro Gly Ala Pro Gly Pro Ala Gly Ala
145                 150                 155                 160

Arg Gly Asn Asp Gly Ala Thr Gly Ala Ala Gly Pro Pro Gly Pro Thr
                165                 170                 175

Gly Pro Ala Gly Pro Pro Gly Phe Pro Gly Ala Val Gly Ala Lys Gly
            180                 185                 190

Glu Ala Gly Pro Gln Gly Pro Arg Gly Ser Glu Gly Pro Gln Gly Val
        195                 200                 205

Arg Gly Glu Pro Gly Pro Pro Gly Pro Ala Gly Ala Ala Gly Pro Ala
    210                 215                 220

Gly Asn Pro Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Ala Asn Gly
225                 230                 235                 240

Ala Pro Gly Ile Ala Gly Ala Pro Gly Phe Pro Gly Ala Arg Gly Pro
                245                 250                 255

Ser Gly Pro Gln Gly Pro Gly Gly Pro Pro Gly Pro Lys Gly Asn Ser
            260                 265                 270

Gly Glu Pro Gly Ala Pro Gly Ser Lys Gly Asp Thr Gly Ala Lys Gly
```

275 280 285

Glu Pro Gly Pro Val Gly Val Gln Gly Pro Pro Gly Pro Ala Gly Glu
290 295 300

Glu Gly Lys Arg Gly Ala Arg Gly Glu Pro Gly Pro Thr Gly Leu Pro
305 310 315 320

Gly Pro Pro Gly Glu Arg Gly Gly Pro Gly Ser Arg Gly Phe Pro Gly
325 330 335

Ala Asp Gly Val Ala Gly Pro Lys Gly Pro Ala Gly Glu Arg Gly Ser
340 345 350

Pro Gly Pro Ala Gly Pro Lys Gly Ser Pro Gly Glu Ala Gly Arg Pro
355 360 365

Gly Glu Ala Gly Leu Pro Gly Ala Lys Gly Leu Thr Gly Ser Pro Gly
370 375 380

Ser Pro Gly Pro Asp Gly Lys Thr Gly Pro Pro Gly Pro Ala Gly Gln
385 390 395 400

Asp Gly Arg Pro Gly Pro Pro Gly Pro Pro Gly Ala Arg Gly Gln Ala
405 410 415

Gly Val Met Gly Phe Pro Gly Pro Lys Gly Ala Ala Gly Glu Pro Gly
420 425 430

Lys Ala Gly Glu Arg Gly Val Pro Gly Pro Pro Gly Ala Val Gly Pro
435 440 445

Ala Gly Lys Asp Gly Glu Ala Gly Ala Gln Gly Pro Pro Gly Pro Ala
450 455 460

Gly Pro Ala Gly Glu Arg Gly Glu Gln Gly Pro Ala Gly Ser Pro Gly
465 470 475 480

Phe Gln Gly Leu Pro Gly Pro Ala Gly Pro Pro Gly Glu Ala Gly Lys
485 490 495

Pro Gly Glu Gln Gly Val Pro Gly Asp Leu Gly Ala Pro Gly Pro Ser
500 505 510

Gly Ala Arg Gly Glu Arg Gly Phe Pro Gly Glu Arg Gly Val Gln Gly
515 520 525

Pro Pro Gly Pro Ala Gly Pro Arg Gly Ala Asn Gly Ala Pro Gly Asn
530 535 540

Asp Gly Ala Lys Gly Asp Ala Gly Ala Pro Gly Ala Pro Gly Ser Gln
545 550 555 560

Gly Ala Pro Gly Leu Gln Gly Met Pro Gly Glu Arg Gly Ala Ala Gly
565 570 575

Leu Pro Gly Pro Lys Gly Asp Arg Gly Asp Ala Gly Pro Lys Gly Ala
580 585 590

Asp Gly Ser Pro Gly Lys Asp Gly Val Arg Gly Leu Thr Gly Pro Ile
595 600 605

Gly Pro Pro Gly Pro Ala Gly Ala Pro Gly Asp Lys Gly Glu Ser Gly
610 615 620

Pro Ser Gly Pro Ala Gly Pro Thr Gly Ala Arg Gly Ala Pro Gly Asp

625 630 635 640

Arg Gly Glu Pro Gly Pro Pro Gly Pro Ala Gly Phe Ala Gly Pro Pro
645 650 655

Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Glu Pro Gly Asp Ala Gly
660 665 670

Ala Lys Gly Asp Ala Gly Pro Pro Gly Pro Ala Gly Pro Ala Gly Pro
675 680 685

Pro Gly Pro Ile Gly Asn Val Gly Ala Pro Gly Ala Lys Gly Ala Arg
690 695 700

Gly Ser Ala Gly Pro Pro Gly Ala Thr Gly Phe Pro Gly Ala Ala Gly
705 710 715 720

Arg Val Gly Pro Pro Gly Pro Ser Gly Asn Ala Gly Pro Pro Gly Pro
725 730 735

Pro Gly Pro Ala Gly Lys Glu Gly Gly Lys Gly Pro Arg Gly Glu Thr
740 745 750

Gly Pro Ala Gly Arg Pro Gly Glu Val Gly Pro Pro Gly Pro Pro Gly
755 760 765

Pro Ala Gly Glu Lys Gly Ser Pro Gly Ala Asp Gly Pro Ala Gly Ala
770 775 780

Pro Gly Thr Pro Gly Pro Gln Gly Ile Ala Gly Gln Arg Gly Val Val
785 790 795 800

Gly Leu Pro Gly Gln Arg Gly Glu Arg Gly Phe Pro Gly Leu Pro Gly
805 810 815

Pro Ser Gly Glu Pro Gly Lys Gln Gly Pro Ser Gly Ala Ser Gly Glu
820 825 830

Arg Gly Pro Pro Gly Pro Met Gly Pro Pro Gly Leu Ala Gly Pro Pro
835 840 845

Gly Glu Ser Gly Arg Glu Gly Ala Pro Ala Ala Glu Gly Ser Pro Gly
850 855 860

Arg Asp Gly Ser Pro Gly Ala Lys Gly Asp Arg Gly Glu Thr Gly Pro
865 870 875 880

Ala Gly Pro Pro Gly Ala Xaa Gly Ala Xaa Gly Ala Pro Gly Pro Val
885 890 895

Gly Pro Ala Gly Lys Ser Gly Asp Arg Gly Glu Thr Gly Pro Ala Gly
900 905 910

Pro Ala Gly Pro Val Gly Pro Ala Gly Ala Arg Gly Pro Ala Gly Pro
915 920 925

Gln Gly Pro Arg Gly Asp Lys Gly Glu Thr Gly Glu Gln Gly Asp Arg
930 935 940

Gly Ile Lys Gly His Arg Gly Phe Ser Gly Leu Gln Gly Pro Pro Gly
945 950 955 960

Pro Pro Gly Ser Pro Gly Glu Gln Gly Pro Ser Gly Ala Ser Gly Pro
965 970 975

Ala Gly Pro Arg Gly Pro Pro Gly Ser Ala Gly Ala Pro Gly Lys Asp

980 985 990

Gly Leu Asn Gly Leu Pro Gly Pro Ile Gly Pro Pro Gly Pro Arg Gly
995 1000 1005

Arg Thr Gly Asp Ala Gly Pro Val Gly Pro Pro Gly Pro Pro Gly Pro
1010 1015 1020

Pro Gly Pro Pro Gly Pro Pro Ser Ala Gly Phe Asp Phe Ser Phe Leu
1025 1030 1035 1040

Pro Gln Pro Pro Gln Glu Lys Ala His Asp Gly Gly Arg Tyr Tyr Arg
1045 1050 1055

Ala Arg Ser Gln Arg Ala Arg Lys Lys Asn Lys Asn Cys Arg Arg His
1060 1065 1070

Ser Leu Tyr Val Asp Phe Ser Asp Val Gly Trp Asn Asp Trp Ile Val
1075 1080 1085

Ala Pro Pro Gly Tyr Gln Ala Phe Tyr Cys His Gly Asp Cys Pro Phe
1090 1095 1100

Pro Leu Ala Asp His Leu Asn Ser Thr Asn His Ala Ile Val Gln Thr
1105 1110 1115 1120

Leu Val Asn Ser Val Asn Ser Ser Ile Pro Lys Ala Cys Cys Val Pro
1125 1130 1135

Thr Glu Leu Ser Ala Ile Ser Met Leu Tyr Leu Asp Glu Tyr Asp Lys
1140 1145 1150

Val Val Leu Lys Asn Tyr Gln Glu Met Val Val Glu Gly Cys Gly Cys
1155 1160 1165

Arg

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1171 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Gln Leu Ser Tyr Gly Tyr Asp Glu Lys Ser Thr Gly Gly Ile Ser Val
1 5 10 15

Pro Gly Pro Met Gly Pro Ser Gly Pro Arg Gly Leu Pro Gly Pro Pro
20 25 30

Gly Ala Pro Gly Pro Gln Gly Phe Gln Gly Pro Pro Gly Glu Pro Gly
35 40 45

Glu Pro Gly Ala Ser Gly Pro Met Gly Pro Arg Gly Pro Pro Gly Pro
50 55 60

Pro Gly Lys Asn Gly Asp Asp Gly Glu Ala Gly Lys Pro Gly Arg Pro
65 70 75 80

Gly Glu Arg Gly Pro Pro Gly Pro Gln Gly Ala Arg Gly Leu Pro Gly
85 90 95

```
Thr Ala Gly Leu Pro Gly Met Lys Gly His Arg Gly Phe Ser Gly Leu
            100                 105                 110

Asp Gly Ala Lys Gly Asp Ala Gly Pro Ala Gly Pro Lys Gly Glu Pro
        115                 120                 125

Gly Ser Pro Gly Glu Asn Gly Ala Pro Gly Gln Met Gly Pro Arg Gly
    130                 135                 140

Leu Pro Gly Glu Arg Gly Arg Pro Gly Ala Pro Gly Pro Ala Gly Ala
145                 150                 155                 160

Arg Gly Asn Asp Gly Ala Thr Gly Ala Ala Gly Pro Pro Gly Pro Thr
                165                 170                 175

Gly Pro Ala Gly Pro Pro Gly Phe Pro Gly Ala Val Gly Ala Lys Gly
            180                 185                 190

Glu Ala Gly Pro Gln Gly Pro Arg Gly Ser Glu Gly Pro Gln Gly Val
        195                 200                 205

Arg Gly Glu Pro Gly Pro Pro Gly Pro Ala Gly Ala Ala Gly Pro Ala
    210                 215                 220

Gly Asn Pro Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Ala Asn Gly
225                 230                 235                 240

Ala Pro Gly Ile Ala Gly Ala Pro Gly Phe Pro Gly Ala Arg Gly Pro
                245                 250                 255

Ser Gly Pro Gln Gly Pro Gly Gly Pro Pro Gly Pro Lys Gly Asn Ser
            260                 265                 270

Gly Glu Pro Gly Ala Pro Gly Ser Lys Gly Asp Thr Gly Ala Lys Gly
        275                 280                 285

Glu Pro Gly Pro Val Gly Val Gln Gly Pro Pro Gly Pro Ala Gly Glu
    290                 295                 300

Glu Gly Lys Arg Gly Ala Arg Gly Glu Pro Gly Pro Thr Gly Leu Pro
305                 310                 315                 320

Gly Pro Pro Gly Glu Arg Gly Gly Pro Gly Ser Arg Gly Phe Pro Gly
                325                 330                 335

Ala Asp Gly Val Ala Gly Pro Lys Gly Pro Ala Gly Glu Arg Gly Ser
            340                 345                 350

Pro Gly Pro Ala Gly Pro Lys Gly Ser Pro Gly Glu Ala Gly Arg Pro
        355                 360                 365

Gly Glu Ala Gly Leu Pro Gly Ala Lys Gly Leu Thr Gly Ser Pro Gly
    370                 375                 380

Ser Pro Gly Pro Asp Gly Lys Thr Gly Pro Pro Gly Pro Ala Gly Gln
385                 390                 395                 400

Asp Gly Arg Pro Gly Pro Pro Gly Pro Pro Gly Ala Arg Gly Gln Ala
                405                 410                 415

Gly Val Met Gly Phe Pro Gly Pro Lys Gly Ala Ala Gly Glu Pro Gly
            420                 425                 430

Lys Ala Gly Glu Arg Gly Val Pro Gly Pro Pro Gly Ala Val Gly Pro
        435                 440                 445
```

Ala Gly Lys Asp Gly Glu Ala Gly Ala Gln Gly Pro Pro Gly Pro Ala
450 455 460

Gly Pro Ala Gly Glu Arg Gly Glu Gln Gly Pro Ala Gly Ser Pro Gly
465 470 475 480

Phe Gln Gly Leu Pro Gly Pro Ala Gly Pro Pro Gly Glu Ala Gly Lys
485 490 495

Pro Gly Glu Gln Gly Val Pro Gly Asp Leu Gly Ala Pro Gly Pro Ser
500 505 510

Gly Ala Arg Gly Glu Arg Gly Phe Pro Gly Glu Arg Gly Val Gln Gly
515 520 525

Pro Pro Gly Pro Ala Gly Pro Arg Gly Ala Asn Gly Ala Pro Gly Asn
530 535 540

Asp Gly Ala Lys Gly Asp Ala Gly Ala Pro Gly Ala Pro Gly Ser Gln
545 550 555 560

Gly Ala Pro Gly Leu Gln Gly Met Pro Gly Glu Arg Gly Ala Ala Gly
565 570 575

Leu Pro Gly Pro Lys Gly Asp Arg Gly Asp Ala Gly Pro Lys Gly Ala
580 585 590

Asp Gly Ser Pro Gly Lys Asp Gly Val Arg Gly Leu Thr Gly Pro Ile
595 600 605

Gly Pro Pro Gly Pro Ala Gly Ala Pro Gly Asp Lys Gly Glu Ser Gly
610 615 620

Pro Ser Gly Pro Ala Gly Pro Thr Gly Ala Arg Gly Ala Pro Gly Asp
625 630 635 640

Arg Gly Glu Pro Gly Pro Pro Gly Pro Ala Gly Phe Ala Gly Pro Pro
645 650 655

Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Glu Pro Gly Asp Ala Gly
660 665 670

Ala Lys Gly Asp Ala Gly Pro Pro Gly Pro Ala Gly Pro Ala Gly Pro
675 680 685

Pro Gly Pro Ile Gly Asn Val Gly Ala Pro Gly Ala Lys Gly Ala Arg
690 695 700

Gly Ser Ala Gly Pro Pro Gly Ala Thr Gly Phe Pro Gly Ala Ala Gly
705 710 715 720

Arg Val Gly Pro Pro Gly Pro Ser Gly Asn Ala Gly Pro Pro Gly Pro
725 730 735

Pro Gly Pro Ala Gly Lys Glu Gly Gly Lys Gly Pro Arg Gly Glu Thr
740 745 750

Gly Pro Ala Gly Arg Pro Gly Glu Val Gly Pro Pro Gly Pro Pro Gly
755 760 765

Pro Ala Gly Glu Lys Gly Ser Pro Gly Ala Asp Gly Pro Ala Gly Ala
770 775 780

Pro Gly Thr Pro Gly Pro Gln Gly Ile Ala Gly Gln Arg Gly Val Val
785 790 795 800

```
Gly Leu Pro Gly Gln Arg Gly Glu Arg Gly Phe Pro Gly Leu Pro Gly
                805                 810                 815

Pro Ser Gly Glu Pro Gly Lys Gln Gly Pro Ser Gly Ala Ser Gly Glu
            820                 825                 830

Arg Gly Pro Pro Gly Pro Met Gly Pro Pro Gly Leu Ala Gly Pro Pro
        835                 840                 845

Gly Glu Ser Gly Arg Glu Gly Ala Pro Ala Ala Glu Gly Ser Pro Gly
    850                 855                 860

Arg Asp Gly Ser Pro Gly Ala Lys Gly Asp Arg Gly Glu Thr Gly Pro
865                 870                 875                 880

Ala Gly Pro Pro Gly Ala Xaa Gly Ala Xaa Gly Ala Pro Gly Pro Val
                885                 890                 895

Gly Pro Ala Gly Lys Ser Gly Asp Arg Gly Glu Thr Gly Pro Ala Gly
            900                 905                 910

Pro Ala Gly Pro Val Gly Pro Ala Gly Ala Arg Gly Pro Ala Gly Pro
        915                 920                 925

Gln Gly Pro Arg Gly Asp Lys Gly Glu Thr Gly Glu Gln Gly Asp Arg
    930                 935                 940

Gly Ile Lys Gly His Arg Gly Phe Ser Gly Leu Gln Gly Pro Pro Gly
945                 950                 955                 960

Pro Pro Gly Ser Pro Gly Glu Gln Gly Pro Ser Gly Ala Ser Gly Pro
                965                 970                 975

Ala Gly Pro Arg Gly Pro Pro Gly Ser Ala Gly Ala Pro Gly Lys Asp
            980                 985                 990

Gly Leu Asn Gly Leu Pro Gly Pro Ile Gly Pro Pro Gly Pro Arg Gly
        995                1000                1005

Arg Thr Gly Asp Ala Gly Pro Val Gly Pro Pro Gly Pro Pro Gly Pro
    1010                1015                1020

Pro Gly Pro Pro Gly Pro Pro Ser Ala Gly Phe Asp Phe Ser Phe Leu
1025                1030                1035                1040

Pro Gln Pro Pro Gln Glu Lys Ala His Asp Gly Gly Arg Tyr Tyr Arg
                1045                1050                1055

Ala Arg Ser Ala Leu Asp Thr Asn Tyr Cys Phe Ser Ser Thr Glu Lys
            1060                1065                1070

Asn Cys Cys Val Arg Gln Leu Tyr Ile Asp Phe Arg Lys Asp Leu Gly
        1075                1080                1085

Trp Lys Trp Ile His Glu Pro Lys Gly Tyr His Ala Asn Phe Cys Leu
    1090                1095                1100

Gly Pro Cys Pro Tyr Ile Trp Ser Leu Asp Thr Gln Tyr Ser Lys Val
1105                1110                1115                1120

Leu Ala Leu Tyr Asn Gln His Asn Pro Gly Ala Ser Ala Ala Pro Cys
                1125                1130                1135

Cys Val Pro Gln Ala Leu Glu Pro Leu Pro Ile Val Tyr Tyr Val Gly
            1140                1145                1150
```

```
Arg Lys Pro Lys Val Glu Gln Leu Ser Asn Met Ile Val Arg Ser Cys
        1155                1160                1165

Lys Cys Ser .
    1170
```

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1388 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

```
Gln Leu Ser Tyr Gly Tyr Asp Glu Lys Ser Thr Gly Gly Ile Ser Val
 1               5                  10                  15

Pro Gly Pro Met Gly Pro Ser Gly Pro Arg Gly Leu Pro Gly Pro Pro
            20                  25                  30

Gly Ala Pro Gly Pro Gln Gly Phe Gln Gly Pro Pro Gly Glu Pro Gly
        35                  40                  45

Glu Pro Gly Ala Ser Gly Pro Met Gly Pro Arg Gly Pro Pro Gly Pro
    50                  55                  60

Pro Gly Lys Asn Gly Asp Asp Gly Glu Ala Gly Lys Pro Gly Arg Pro
65                  70                  75                  80

Gly Glu Arg Gly Pro Pro Gly Pro Gln Gly Ala Arg Gly Leu Pro Gly
                85                  90                  95

Thr Ala Gly Leu Pro Gly Met Lys Gly His Arg Gly Phe Ser Gly Leu
            100                 105                 110

Asp Gly Ala Lys Gly Asp Ala Gly Pro Ala Gly Pro Lys Gly Glu Pro
        115                 120                 125

Gly Ser Pro Gly Glu Asn Gly Ala Pro Gly Gln Met Gly Pro Arg Gly
    130                 135                 140

Leu Pro Gly Glu Arg Gly Arg Pro Gly Ala Pro Gly Pro Ala Gly Ala
145                 150                 155                 160

Arg Gly Asn Asp Gly Ala Thr Gly Ala Ala Gly Pro Pro Gly Pro Thr
                165                 170                 175

Gly Pro Ala Gly Pro Pro Gly Phe Pro Gly Ala Val Gly Ala Lys Gly
            180                 185                 190

Glu Ala Gly Pro Gln Gly Pro Arg Gly Ser Glu Gly Pro Gln Gly Val
        195                 200                 205

Arg Gly Glu Pro Gly Pro Pro Gly Pro Ala Gly Ala Ala Gly Pro Ala
    210                 215                 220

Gly Asn Pro Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Ala Asn Gly
225                 230                 235                 240

Ala Pro Gly Ile Ala Gly Ala Pro Gly Phe Pro Gly Ala Arg Gly Pro
                245                 250                 255
```

Ser Gly Pro Gln Gly Pro Gly Gly Pro Pro Gly Pro Lys Gly Asn Ser
260 265 270

Gly Glu Pro Gly Ala Pro Gly Ser Lys Gly Asp Thr Gly Ala Lys Gly
275 280 285

Glu Pro Gly Pro Val Gly Val Gln Gly Pro Pro Gly Pro Ala Gly Glu
290 295 300

Glu Gly Lys Arg Gly Ala Arg Gly Glu Pro Gly Pro Thr Gly Leu Pro
305 310 315 320

Gly Pro Pro Gly Glu Arg Gly Gly Pro Gly Ser Arg Gly Phe Pro Gly
325 330 335

Ala Asp Gly Val Ala Gly Pro Lys Gly Pro Ala Gly Glu Arg Gly Ser
340 345 350

Pro Gly Pro Ala Gly Pro Lys Gly Ser Pro Gly Glu Ala Gly Arg Pro
355 360 365

Gly Glu Ala Gly Leu Pro Gly Ala Lys Gly Leu Thr Gly Ser Pro Gly
370 375 380

Ser Pro Gly Pro Asp Gly Lys Thr Gly Pro Pro Gly Pro Ala Gly Gln
385 390 395 400

Asp Gly Arg Pro Gly Pro Pro Gly Pro Pro Gly Ala Arg Gly Gln Ala
405 410 415

Gly Val Met Gly Phe Pro Gly Pro Lys Gly Ala Ala Gly Glu Pro Gly
420 425 430

Lys Ala Gly Glu Arg Gly Val Pro Gly Pro Pro Gly Ala Val Gly Pro
435 440 445

Ala Gly Lys Asp Gly Glu Ala Gly Ala Gln Gly Pro Pro Gly Pro Ala
450 455 460

Gly Pro Ala Gly Glu Arg Gly Glu Gln Gly Pro Ala Gly Ser Pro Gly
465 470 475 480

Phe Gln Gly Leu Pro Gly Pro Ala Gly Pro Pro Gly Glu Ala Gly Lys
485 490 495

Pro Gly Glu Gln Gly Val Pro Gly Asp Leu Gly Ala Pro Gly Pro Ser
500 505 510

Gly Ala Arg Gly Glu Arg Gly Phe Pro Gly Glu Arg Gly Val Gln Gly
515 520 525

Pro Pro Gly Pro Ala Gly Pro Arg Gly Ala Asn Gly Ala Pro Gly Asn
530 535 540

Asp Gly Ala Lys Gly Asp Ala Gly Ala Pro Gly Ala Pro Gly Ser Gln
545 550 555 560

Gly Ala Pro Gly Leu Gln Gly Met Pro Gly Glu Arg Gly Ala Ala Gly
565 570 575

Leu Pro Gly Pro Lys Gly Asp Arg Gly Asp Ala Gly Pro Lys Gly Ala
580 585 590

Asp Gly Ser Pro Gly Lys Asp Gly Val Arg Gly Leu Thr Gly Pro Ile
595 600 605

Gly Pro Pro Gly Pro Ala Gly Ala Pro Gly Asp Lys Gly Glu Ser Gly
610 615 620

Pro Ser Gly Pro Ala Gly Pro Thr Gly Ala Arg Gly Ala Pro Gly Asp
625 630 635 640

Arg Gly Glu Pro Gly Pro Pro Gly Pro Ala Gly Phe Ala Gly Pro Pro
645 650 655

Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Glu Pro Gly Asp Ala Gly
660 665 670

Ala Lys Gly Asp Ala Gly Pro Pro Gly Pro Ala Gly Pro Ala Gly Pro
675 680 685

Pro Gly Pro Ile Gly Asn Val Gly Ala Pro Gly Ala Lys Gly Ala Arg
690 695 700

Gly Ser Ala Gly Pro Pro Gly Ala Thr Gly Phe Pro Gly Ala Ala Gly
705 710 715 720

Arg Val Gly Pro Pro Gly Pro Ser Gly Asn Ala Gly Pro Pro Gly Pro
725 730 735

Pro Gly Pro Ala Gly Lys Glu Gly Gly Lys Gly Pro Arg Gly Glu Thr
740 745 750

Gly Pro Ala Gly Arg Pro Gly Glu Val Gly Pro Pro Gly Pro Pro Gly
755 760 765

Pro Ala Gly Glu Lys Gly Ser Pro Gly Ala Asp Gly Pro Ala Gly Ala
770 775 780

Pro Gly Thr Pro Gly Pro Gln Gly Ile Ala Gly Gln Arg Gly Val Val
785 790 795 800

Gly Leu Pro Gly Gln Arg Gly Glu Arg Gly Phe Pro Gly Leu Pro Gly
805 810 815

Pro Ser Gly Glu Pro Gly Lys Gln Gly Pro Ser Gly Ala Ser Gly Glu
820 825 830

Arg Gly Pro Pro Gly Pro Met Gly Pro Pro Gly Leu Ala Gly Pro Pro
835 840 845

Gly Glu Ser Gly Arg Glu Gly Ala Pro Ala Ala Glu Gly Ser Pro Gly
850 855 860

Arg Asp Gly Ser Pro Gly Ala Lys Gly Asp Arg Gly Glu Thr Gly Pro
865 870 875 880

Ala Gly Pro Pro Gly Ala Xaa Gly Ala Xaa Gly Ala Pro Gly Pro Val
885 890 895

Gly Pro Ala Gly Lys Ser Gly Asp Arg Gly Glu Thr Gly Pro Ala Gly
900 905 910

Pro Ala Gly Pro Val Gly Pro Ala Gly Ala Arg Gly Pro Ala Gly Pro
915 920 925

Gln Gly Pro Arg Gly Asp Lys Gly Glu Thr Gly Glu Gln Gly Asp Arg
930 935 940

Gly Ile Lys Gly His Arg Gly Phe Ser Gly Leu Gln Gly Pro Pro Gly
945 950 955 960

```
Pro Pro Gly Ser Pro Gly Glu Gln Gly Pro Ser Gly Ala Ser Gly Pro
                965                 970                 975

Ala Gly Pro Arg Gly Pro Pro Gly Ser Ala Gly Ala Pro Gly Lys Asp
            980                 985                 990

Gly Leu Asn Gly Leu Pro Gly Pro Ile Gly Pro Pro Gly Pro Arg Gly
        995                 1000                1005

Arg Thr Gly Asp Ala Gly Pro Val Gly Pro Pro Gly Pro Pro Gly Pro
    1010                1015                1020

Pro Gly Pro Pro Gly Pro Pro Ser Ala Gly Phe Asp Phe Ser Phe Leu
1025                1030                1035                1040

Pro Gln Pro Pro Gln Glu Lys Ala His Asp Arg Gly Arg Tyr Tyr Arg
                1045                1050                1055

Ala Arg Ser Asp Glu Ala Ser Gly Ile Gly Pro Glu Val Pro Asp Asp
            1060                1065                1070

Arg Asp Phe Glu Pro Ser Leu Gly Pro Val Cys Pro Phe Arg Cys Gln
        1075                1080                1085

Cys His Leu Arg Val Val Gln Cys Ser Asp Leu Gly Leu Asp Lys Val
    1090                1095                1100

Pro Lys Asp Leu Pro Pro Asp Thr Thr Leu Leu Asp Leu Gln Asn Asn
1105                1110                1115                1120

Lys Ile Thr Glu Ile Lys Asp Gly Asp Phe Lys Asn Leu Lys Asn Leu
                1125                1130                1135

His Ala Leu Ile Leu Val Asn Asn Lys Ile Ser Lys Val Ser Pro Gly
            1140                1145                1150

Ala Phe Thr Pro Leu Val Lys Leu Glu Arg Leu Tyr Leu Ser Lys Asn
        1155                1160                1165

Gln Leu Lys Glu Leu Pro Glu Lys Met Pro Lys Thr Leu Gln Glu Leu
    1170                1175                1180

Arg Ala His Glu Asn Glu Ile Thr Lys Val Arg Lys Val Thr Phe Asn
1185                1190                1195                1200

Gly Leu Asn Gln Met Ile Val Ile Glu Leu Gly Thr Asn Pro Leu Lys
                1205                1210                1215

Ser Ser Gly Ile Glu Asn Gly Ala Phe Gln Gly Met Lys Lys Leu Ser
            1220                1225                1230

Tyr Ile Arg Ile Ala Asp Thr Asn Ile Thr Ser Ile Pro Gln Gly Leu
        1235                1240                1245

Pro Pro Ser Leu Thr Glu Leu His Leu Asp Gly Asn Lys Ile Ser Arg
    1250                1255                1260

Val Asp Ala Ala Ser Leu Lys Gly Leu Asn Asn Leu Ala Lys Leu Gly
1265                1270                1275                1280

Leu Ser Phe Asn Ser Ile Ser Ala Val Asp Asn Gly Ser Leu Ala Asn
                1285                1290                1295

Thr Pro His Leu Arg Glu Leu His Leu Asp Asn Asn Lys Leu Thr Arg
            1300                1305                1310
```

```
Val Pro Gly Gly Leu Ala Glu His Lys Tyr Ile Gln Val Val Tyr Leu
        1315                1320                1325

His Asn Asn Asn Ile Ser Val Val Gly Ser Ser Asp Phe Cys Pro Pro
    1330                1335                1340

Gly His Asn Thr Lys Lys Ala Ser Tyr Ser Gly Val Ser Leu Phe Ser
1345                1350                1355                1360

Asn Pro Val Gln Tyr Trp Glu Ile Gln Pro Ser Thr Phe Arg Cys Val
                1365                1370                1375

Tyr Val Arg Ser Ala Ile Gln Leu Gly Asn Tyr Lys
            1380                1385
```

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1107 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

```
Gln Leu Ser Tyr Gly Tyr Asp Glu Lys Ser Thr Gly Gly Ile Ser Val
  1               5                  10                  15

Pro Gly Pro Met Gly Pro Ser Gly Pro Arg Gly Leu Pro Gly Pro Pro
            20                  25                  30

Gly Ala Pro Gly Pro Gln Gly Phe Gln Gly Pro Pro Gly Glu Pro Gly
        35                  40                  45

Glu Pro Gly Ala Ser Gly Pro Met Gly Pro Arg Gly Pro Pro Gly Pro
    50                  55                  60

Pro Gly Lys Asn Gly Asp Asp Gly Glu Ala Gly Lys Pro Gly Arg Pro
 65                  70                  75                  80

Gly Glu Arg Gly Pro Pro Gly Pro Gln Gly Ala Arg Gly Leu Pro Gly
                85                  90                  95

Thr Ala Gly Leu Pro Gly Met Lys Gly His Arg Gly Phe Ser Gly Leu
            100                 105                 110

Asp Gly Ala Lys Gly Asp Ala Gly Pro Ala Gly Pro Lys Gly Glu Pro
        115                 120                 125

Gly Ser Pro Gly Glu Asn Gly Ala Pro Gly Gln Met Gly Pro Arg Gly
    130                 135                 140

Leu Pro Gly Glu Arg Gly Arg Pro Gly Ala Pro Gly Pro Ala Gly Ala
145                 150                 155                 160

Arg Gly Asn Asp Gly Ala Thr Gly Ala Ala Gly Pro Pro Gly Pro Thr
                165                 170                 175

Gly Pro Ala Gly Pro Pro Gly Phe Pro Gly Ala Val Gly Ala Lys Gly
            180                 185                 190

Glu Ala Gly Pro Gln Gly Pro Arg Gly Ser Glu Gly Pro Gln Gly Val
        195                 200                 205

Arg Gly Glu Pro Gly Pro Pro Gly Pro Ala Gly Ala Ala Gly Pro Ala
```

```
    210                 215                 220

Gly Asn Pro Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Ala Asn Gly
225                 230                 235                 240

Ala Pro Gly Ile Ala Gly Ala Pro Gly Phe Pro Gly Ala Arg Gly Pro
                245                 250                 255

Ser Gly Pro Gln Gly Pro Gly Gly Pro Pro Gly Pro Lys Gly Asn Ser
            260                 265                 270

Gly Glu Pro Gly Ala Pro Gly Ser Lys Gly Asp Thr Gly Ala Lys Gly
        275                 280                 285

Glu Pro Gly Pro Val Gly Val Gln Gly Pro Pro Gly Pro Ala Gly Glu
    290                 295                 300

Glu Gly Lys Arg Gly Ala Arg Gly Glu Pro Gly Pro Thr Gly Leu Pro
305                 310                 315                 320

Gly Pro Pro Gly Glu Arg Gly Gly Pro Gly Ser Arg Gly Phe Pro Gly
                325                 330                 335

Ala Asp Gly Val Ala Gly Pro Lys Gly Pro Ala Gly Glu Arg Gly Ser
            340                 345                 350

Pro Gly Pro Ala Gly Pro Lys Gly Ser Pro Gly Glu Ala Gly Arg Pro
        355                 360                 365

Gly Glu Ala Gly Leu Pro Gly Ala Lys Gly Leu Thr Gly Ser Pro Gly
    370                 375                 380

Ser Pro Gly Pro Asp Gly Lys Thr Gly Pro Pro Gly Pro Ala Gly Gln
385                 390                 395                 400

Asp Gly Arg Pro Gly Pro Pro Gly Pro Pro Gly Ala Arg Gly Gln Ala
                405                 410                 415

Gly Val Met Gly Phe Pro Gly Pro Lys Gly Ala Ala Gly Glu Pro Gly
            420                 425                 430

Lys Ala Gly Glu Arg Gly Val Pro Gly Pro Pro Gly Ala Val Gly Pro
        435                 440                 445

Ala Gly Lys Asp Gly Glu Ala Gly Ala Gln Gly Pro Pro Gly Pro Ala
    450                 455                 460

Gly Pro Ala Gly Glu Arg Gly Glu Gln Gly Pro Ala Gly Ser Pro Gly
465                 470                 475                 480

Phe Gln Gly Leu Pro Gly Pro Ala Gly Pro Pro Gly Glu Ala Gly Lys
                485                 490                 495

Pro Gly Glu Gln Gly Val Pro Gly Asp Leu Gly Ala Pro Gly Pro Ser
            500                 505                 510

Gly Ala Arg Gly Glu Arg Gly Phe Pro Gly Glu Arg Gly Val Gln Gly
        515                 520                 525

Pro Pro Gly Pro Ala Gly Pro Arg Gly Ala Asn Gly Ala Pro Gly Asn
    530                 535                 540

Asp Gly Ala Lys Gly Asp Ala Gly Ala Pro Gly Ala Pro Gly Ser Gln
545                 550                 555                 560

Gly Ala Pro Gly Leu Gln Gly Met Pro Gly Glu Arg Gly Ala Ala Gly
```

```
                        565                 570                 575

Leu Pro Gly Pro Lys Gly Asp Arg Gly Asp Ala Gly Pro Lys Gly Ala
            580                 585                 590

Asp Gly Ser Pro Gly Lys Asp Gly Val Arg Gly Leu Thr Gly Pro Ile
        595                 600                 605

Gly Pro Pro Gly Pro Ala Gly Ala Pro Gly Asp Lys Gly Glu Ser Gly
    610                 615                 620

Pro Ser Gly Pro Ala Gly Pro Thr Gly Ala Arg Gly Ala Pro Gly Asp
625                 630                 635                 640

Arg Gly Glu Pro Gly Pro Pro Gly Pro Ala Gly Phe Ala Gly Pro Pro
                645                 650                 655

Gly Ala Asp Gly Gln Pro Gly Ala Lys Gly Glu Pro Gly Asp Ala Gly
            660                 665                 670

Ala Lys Gly Asp Ala Gly Pro Pro Gly Pro Ala Gly Pro Ala Gly Pro
        675                 680                 685

Pro Gly Pro Ile Gly Asn Val Gly Ala Pro Gly Ala Lys Gly Ala Arg
    690                 695                 700

Gly Ser Ala Gly Pro Pro Gly Ala Thr Gly Phe Pro Gly Ala Ala Gly
705                 710                 715                 720

Arg Val Gly Pro Pro Gly Pro Ser Gly Asn Ala Gly Pro Pro Gly Pro
                725                 730                 735

Pro Gly Pro Ala Gly Lys Glu Gly Gly Lys Gly Pro Arg Gly Glu Thr
            740                 745                 750

Gly Pro Ala Gly Arg Pro Gly Glu Val Gly Pro Pro Gly Pro Pro Gly
        755                 760                 765

Pro Ala Gly Glu Lys Gly Ser Pro Gly Ala Asp Gly Pro Ala Gly Ala
    770                 775                 780

Pro Gly Thr Pro Gly Pro Gln Gly Ile Ala Gly Gln Arg Gly Val Val
785                 790                 795                 800

Gly Leu Pro Gly Gln Arg Gly Glu Arg Gly Phe Pro Gly Leu Pro Gly
                805                 810                 815

Pro Ser Gly Glu Pro Gly Lys Gln Gly Pro Ser Gly Ala Ser Gly Glu
            820                 825                 830

Arg Gly Pro Pro Gly Pro Met Gly Pro Pro Gly Leu Ala Gly Pro Pro
        835                 840                 845

Gly Glu Ser Gly Arg Glu Gly Ala Pro Ala Ala Glu Gly Ser Pro Gly
    850                 855                 860

Arg Asp Gly Ser Pro Gly Ala Lys Gly Asp Arg Gly Glu Thr Gly Pro
865                 870                 875                 880

Ala Gly Pro Pro Gly Ala Xaa Gly Ala Xaa Gly Ala Pro Gly Pro Val
                885                 890                 895

Gly Pro Ala Gly Lys Ser Gly Asp Arg Gly Glu Thr Gly Pro Ala Gly
            900                 905                 910

Pro Ala Gly Pro Val Gly Pro Ala Gly Ala Arg Gly Pro Ala Gly Pro
```

```
        915                     920                     925

Gln Gly Pro Arg Gly Asp Lys Gly Glu Thr Gly Glu Gln Gly Asp Arg
    930                     935                     940

Gly Ile Lys Gly His Arg Gly Phe Ser Gly Leu Gln Gly Pro Pro Gly
945                     950                     955                     960

Pro Pro Gly Ser Pro Gly Glu Gln Gly Pro Ser Gly Ala Ser Gly Pro
                965                     970                     975

Ala Gly Pro Arg Gly Pro Pro Gly Ser Ala Gly Ala Pro Gly Lys Asp
            980                     985                     990

Gly Leu Asn Gly Leu Pro Gly Pro Ile Gly Pro Pro Gly Pro Arg Gly
        995                     1000                    1005

Arg Thr Gly Asp Ala Gly Pro Val Gly Pro Pro Gly Pro Pro Gly Pro
    1010                    1015                    1020

Pro Gly Pro Pro Gly Pro Pro Ser Ala Gly Phe Asp Phe Ser Phe Leu
1025                    1030                    1035                    1040

Pro Gln Pro Pro Gln Glu Lys Ala His Asp Gly Gly Arg Tyr Tyr Arg
                1045                    1050                    1055

Ala Arg Ser Pro Lys Asp Leu Pro Pro Asp Thr Thr Leu Leu Asp Leu
            1060                    1065                    1070

Gln Asn Asn Lys Ile Thr Glu Ile Lys Asp Gly Asp Phe Lys Asn Leu
        1075                    1080                    1085

Lys Asn Leu His Ala Leu Ile Leu Val Asn Asn Lys Ile Ser Lys Val
    1090                    1095                    1100

Ser Pro Gly
1105
```

Claims

1. A chimeric DNA construct comprising a domain derived from a DNA sequence encoding a cellular regulatory factor and a domain derived from a DNA sequence encoding an extracellular matrix protein.

2. A chimeric DNA construct according to claim 1, wherein said extracellular matrix protein is selected from the group consisting of collagen, laminin, fibronectin, elastin and fibrin.

3. A chimeric DNA construct according to claim 1 or 2 wherein said cellular regulatory factor is selected from the group consisting of BMP, TGF-β, and decorin.

4. A chimeric DNA construct according to claim 1 or 2 wherein said cellular regulatory factor is selected from the group consisting of, a BMP fragment, a TGF-β fragment and a decorin peptide.

5. The DNA construct according to claim 3, wherein said BMP protein comprises BMP-2B.

6. A cloning vector comprising a DNA construct according to any one of claims 1 to 5.

7. A cloning vector according to claim 6, wherein said cloning vector is selected from the group consisting of plasmids, phages, cosmids and artificial chromosomes.

8. A cloning vector according to claim 6 or 7, wherein said cloning vector is pMal.

9. A cell transformed by a cloning vector according to any one of claims 6 to 8.

10. A cell according to claim 9 wherein said cell is selected from the group consisting of E. Coli, HeLa, 3T3, CHO, SP2, Sf9, Sf21, and High Five.

11. A chimeric protein comprising a domain derived from a cellular regulatory factor and a domain derived from an extracellular matrix protein.

12. A chimeric protein according to claim 11, wherein said extracellular matrix protein is selected from the group consisting of collagen, fibronectin, elastin, laminin and fibrin.

13. A chimeric protein according to claim 11 or 12, wherein said cellular regulatory factor is selected from the group consisting of BMP, TGF-β, decorin and a decorin peptide.

14. A method of manufacturing a chimeric cellular regulatory factor/extracellular matrix protein comprising: transforming a cell with the vector according to any one of claims 6 to 8; culturing said cell in a suitable culture medium; and obtaining said chimeric cellular regulatory factor/extracellular matrix protein from said culture medium.

15. A pharmaceutical vehicle for delivery of a therapeutically active substance comprising a chimeric protein having at least two domains, wherein one of said domains is at least a portion of an extracellular matrix protein and another of said domains is at least a portion of a therapeutically active moiety and said domains are covalently linked.

16. A pharmaceutical composition comprising a chimeric protein comprising a domain derived from a cellular regulatory factor and a domain derived from an extracellular matrix protein and a pharmaceutically acceptable vehicle.

17. A pharmaceutical composition according to claim 16, wherein said extracellular matrix protein is selected from the group consisting of collagen, fibronectin, elastin and fibrin.

18. An pharmaceutical composition according to claim 16 or 17, wherein said cellular regulatory factor is selected from the group consisting of BMP, TGF-β, decorin and a decorin peptide.

19. A pharmaceutical composition according to any one of claims 16 to 18, wherein said vehicle comprises a material selected from the group consisting of bioabsorbable polymers, bicompatible nonabsorbable polymers, lactoner putty and plaster of Paris.

20. A pharmaceutical composition according to claim 19, wherein said material is selected from the group consisting of lactide, glycolide, trimethylene carbonate, dioxanone, caprolactone, polymethylmethacrylate and hydroxyethyl-methacrylate.

21. A method of preparing a DNA construct comprising: providing DNA which encodes a cellular regulatory factor or fragment thereof;

providing DNA which encodes an extracellular matrix protein or fragment thereof; and operably linking said cellular regulatory factor or fragment thereof encoding DNA to said extracellular matrix protein or fragment thereof encoding DNA to form a chimeric DNA construct.

22. Use of a chimeric protein according to any one of claims 11 to 13 or a pharmaceutical composition according to any one of claims 16 to 20 for the manufacture of a medicament for the prevention or treatment of disease.

23. Use of a chimeric protein according to claim 13 or a pharmaceutical composition according to any one of claims 18 to 20 for the manufacture of a osteogenic agent.

24. Use of a chimeric protein according to claim 13 or a pharmaceutical composition according to any one of claims 18 to 20, wherein said cellular regulatory factor is BMP, for the manufacture of a medicament for inducing bone and/or cartilage formation.

25. Use of a chimeric protein according to claim 13 or a pharmaceutical composition according to any one of claims 18 to 20, wherein said cellular regulatory factor is TGF-β, for the manufacture of a medicament for inducing soft tissue repair.

26. Use of a chimeric protein according to claim 13 or a pharmaceutical composition according to any one of claims 18 to 20, wherein said cellular regulatory factor is decorin or a decorin peptide for the manufacture of a medicament for reducing scar formation.